Metabolic pathway engineering of the toluene degradation pathway.

A METABOLIC PATHWAY ENGINEERING STUDY OF THE TOL META-CLEAVAGE PATHWAY

A thesis submitted to the University of London
for the degree of DOCTOR OF PHILOSOPHY

by

Lucy Regan B.Eng (Hons.), M.Sc.

BIBL.
LONDIN.
UNIV.

The fascination with what's difficult.....

W.B. Yeats

Acknowledgements

I would like to thank my supervisor Peter Dunnill for his support and encouragement, as well as my co-workers at UCL who ensured that my time there was stimulating and enjoyable.
I would especially like to thank my parents for bringing me up to realise the value of education and continued learning.
Also, special thanks to Tom.

ABSTRACT

This thesis addresses the problem of how to examine a metabolic pathway and identify what are the key elements, specifically with respect to rate-limitation. The aim is to be able to analyze a pathway, identify the bottlenecks and implement genetic modifications to remove these bottlenecks. This is done by defining the system of interest and developing a predictive model using kinetic data. The model predictions can then be verified using fermentation data and genetic techniques to make the appropriate changes for improved performance.

The test system chosen for this study was the TOL meta-cleavage pathway for the degradation of benzoate. This system was chosen on the basis of the application of pathway engineering principles to other systems. The modelling strategy and software was developed using principles from metabolic control theory and biochemical systems theory. By applying this to the TOL pathway using kinetic data, the control coefficients for the pathway were obtained as well as the system parameters required for the optimization of the pathway.

The simulated results obtained from this model must be validated by experiment. Errors can arise both from incorrect assumptions in the model and from the fact that the kinetic data taken from individual in vitro experiments may not be applicable to the in vivo system. The effect of the presence of the TOL pathway on the behaviour of E.coli JM107 during fermentation was investigated and the transient concentration data necessary to identify the bottlenecks in the pathway measured.

This data is then used to calculate the flux control coefficients for the TOL pathway. The predictive results were verified by the fermentation data which identified the first two enzymes in the pathway as having significant flux control coefficients.

This final chapter also addresses the issue of flux analysis, that is, the calculation of the fluxes in the system to determine where fluxes to unwanted by-products occur and to indicate points of control. A graphical user interface is used to provide a user-friendly and intuitive means of building and customising metabolic pathways which can then be interfaced with instrumentation to provide on-line flux analysis.

CONTENTS

4.0 Introduction

4.1 Choice of media

4.2 Effect of benzoate and plasmid induction

4.3 Optical density and dry weight correlations

4.4 Results of total organic carbon analysis

4.5 Metabolite concentrations

5. Flux and Metabolic Control Analysis

5.0 Introduction

5.1 Flux analysis

- 5.1.0 Introduction
- 5.1.1 Method
- 5.1.2 Construction of pathway modules
- 5.1.3 Flux analysis of TOL pathway
  - 5.1.3.1 Results
- 5.1.4 Flux analysis of organic acid production
  - 5.1.4.1 Results

5.2 Calculation of Control Coefficients from Experimental Data

- 5.2.1 The two-step system
  - 5.2.1.1 Results
- 5.2.2 The refined system
  - 5.2.2.1 Results

List of Figures

1. INTRODUCTION:

1.0 Introduction

The aim of metabolic pathway engineering is to use genetic engineering techniques to alter the cells activity in a desirable way. This involves highly specific additions or deletions to the cells genome to alter the proteins expressed or change the regulation of their expression. Such changes can have complex effects on the cellular metabolism. There are several reasons for this. The first is the way that the new genetic information is introduced to the cell: it will either be chromosomal or plasmid-borne. Introducing plasmids to a cell places a metabolic burden on the cell and can have other effects which have been studied [Axe and Bailey, 1989]. It is also necessary to take into account any feedback repression which may occur as a result of the over-production of a metabolite. These two effects indicate the importance of considering the effects of any modification on the pathways of the whole microorganism. This requires a systematic approach to analysing the behaviour of metabolic pathways and the effects of perturbations upon them. To be able to engineer a pathway for increased productivity or to make new products requires an extensive knowledge of the mechanisms of the pathway and it is necessary to formulate some kind of model of the system. Models can vary in complexity from being purely phenomenological to being highly mathematically structured but their basic purpose is to explain and predict the action of the system. This chapter will describe how different types of knowledge about systems can be used to engineer them for improved performance. The chapter is divided into two sections; the first describes approaches which do not use control theories, the second gives an account of the control theories which are applicable to metabolic systems.

Sections 1.1.1 and 1.1.2 give examples of where a knowledge-based approach has been used in pathway engineering, both in flux optimization and in the creation of new

pathways. Section 1.1.3 then describes the types of models and expert systems which have been developed to describe metabolic pathways. The simplest of the mathematical models are unstructured models. These are models which do not make use of any biochemical or kinetic data but lump details of the metabolic system into one or more pools which are then treated as black boxes. For example, the Monod equation for cell growth assumes the growth of a cell to be determined by Michaelis-Menten kinetics of a limiting substrate, X and lumps all of the other processes together so that an expression for cell growth μ is

$$\mu = \mu_{max} \cdot \frac{X}{K_X + X} \qquad (1)$$

While this expression can provide a good fit under some conditions, it is not universally true, for example where more than one step affects growth rate. A second approach treats the cell as a black box where only the inputs and outputs of substrates, products and energy need be considered. This approach is described in Section 1.1.3.1.2 where a set of linear conservation equations are derived from the element, electrical charge, energy and Gibbs free energy balances. The chemistry of the process is also used in providing the balance equations. The microorganism is specified only by its elemental composition -for example *Acinetobacter calcoaceticus* is $CH_{1.66}O_{0.46}N_{0.22}$.

This method can be used to check the validity of experimental data, to predict maximum yields and to predict the yields of compounds difficult to measure (Papoutsakis 1984, Papoutsakis and Meyer 1985a,b). It has however some drawbacks, one of these being that it can give no insight into cellular regulation by individual steps in the pathway or any of the other details of intracellular metabolism.

The structured models of the cell described in Section 1.1.3.1.3 make use to a greater or lesser extent of the biochemical details of the cells metabolism. In simple cases this will be the pathway stoichiometry (Aiba and Matsuoka 1984, Holms 1986) whereas, in highly structured models, details of initiation and termination of DNA synthesis (Shu and Schuler 1989), initiation of sporulation (Jeong and Ataai 1990) and transcription and translation (Peretti and Bailey 1986) may be included.

Simple flux models discussed in Section 1.1.3.1.1 can be used to identify a correct metabolic model (Aiba and Matsuoka 1979) or to elucidate regulatory and control mechanisms. They require measured levels of metabolites rather than enzyme kinetic data or data on intermediate or enzyme concentrations. However they yield no information on the regulatory enzymes or metabolites in the pathway or the intermediate metabolite concentrations.

It has been shown that by performing a balance over the inputs and the outputs of the cell, the size of the system to be solved can be reduced if the number of independent balance equations is not at its maximum (Tsai and Lee 1988, Noorman *et al* 1991). If this is not the case, the metabolic pathway approach should be used as it gives more information about the internal mechanisms of the cell.

There have been several studies comparing the approaches where the cell inputs and outputs are balanced with those where information on the metabolic pathways is used (Tsai and Lee 1988, Noorman et al 1991) and it has been shown that the balance method can yield extra information and reduce the size of the system to be solved only when the number of independent balance equations is not at its maximum. If this is not the case, the metabolic pathway approach should be used as it gives more information about the internal mechanisms of the cell.

The highly structured models described in Section 1.1.3.1.3 can yield a large amount of data but require a large number of input parameters some of which may be difficult to obtain. For example a study of sporulation in *B.subtilis* entailed the solution of 17 coupled differential equations (using a Cray supercomputer) requiring 210 operational parameters (Jeong and Ataai 1990). As well as the powerful computing and numerical techniques which are required when solving such systems, the results obtained may be ambiguous and errors not immediately noticeable.

Even if all of the details of a metabolic system were known it is often not desirable to construct a highly structured model. Many of the elements of a system may remain relatively unchanged during a process - this is what allows the use of the pseudo-steady state assumption common to many kinetic studies - and variations in them have very little effect on the overall pathway.

Therefore what is required is to identify the elements of the metabolic system which regulate or control the system. The cell balances its metabolic demands by controlling

the concentration of the elements of the pathways and the rates of conversion or the fluxes through the pathways. The control mechanisms used by the cell can act at several different points. It can be at the transcriptional or translational level so that it is the level of the enzyme in the cell which is controlled or it can be at the reaction level where it is the activity of the enzyme which is altered usually by direct interaction with the modulator. Control may also be exerted upon transport of substrates or products across the cellular membrane, or where there is spatial organisation within the cell as in eucaryotes, across the intracellular membranes. Even if the exact mechanisms of control are not known it is possible to obtain quantitative descriptions of how the control is distributed among the elements of a pathway using the theories of metabolic control reviewed in Section 1.2.

Metabolic control theory was developed on two fronts in the early 70's, one based on the derivation of control theorems from the examination of simple linear systems (Heinrich and Rappoport 1974), the other based on a power-law method of expressing reaction rates. These theories of metabolic control and others developed after them (Savageau 1969a,b,1970) differ in details of how the reaction rates are aggregated and whether a rate law approximation is used implicitly or explicitly but all generate a sensitivity analysis of the flux and metabolite concentrations. The sensitivity or control coefficient gives a measure of the fractional change of a dependent variable caused by a fractional change in an independent variable, i.e.

$$C_y^x = S(x,y) = \frac{dX/X}{dY/Y} = \frac{d\ln X}{d\ln Y} \qquad (2)$$

While optimization methods have been widely applied in chemical processes especially with respect to costing, their use in the biological sector has so far been limited by the uncertainty existing in the mathematical representation of biochemical processes and in the data used in the process models. As new modelling techniques are developed they can be combined with existing optimization methods to provide a rational method of metabolic pathway optimization. A systematic methodology is clearly required because of the potential size and complexity of pathways for pharmaceutical products which include reversible steps and loops.

The need for a rational approach to modelling metabolic systems has become greater with the increasing ability to manipulate the conditions inside microorganisms using genetic techniques. Such metabolic models aid the visualization of complex systems and provide predictive results to direct experimental design. More importantly perhaps is the way in which these techniques can be used in the metabolic pathway engineering of cells. For example, to force a cell to make more of an existing product, to insert novel pathways into a cell or to make novel products. Successful metabolic pathway engineering attempts have produced impressive results such as a yield improvement of over 50% in the production of phenylalanine (Backman et al, 1990), the synthesis of indigo by a *Pseudomonas* strain (Mermod *et al* 1986) and the production of novel antibiotics (Hunter and Baumberg 1989).

An overview of the modelling techniques discussed in this chapter and their inter-relationships is shown in Figure 1.1. This figure highlights the techniques which are to be used in this study in later chapters.

The conclusion of this chapter deals with the choice of a test system for study and the strategy for modelling it. The system chosen was the TOL meta--cleavage pathway. The reasons for using this were the extensive data available on its genetics and enzymology and the availability of genetic systems which could be transformed into *E.coli* which is a particularly well defined organism.

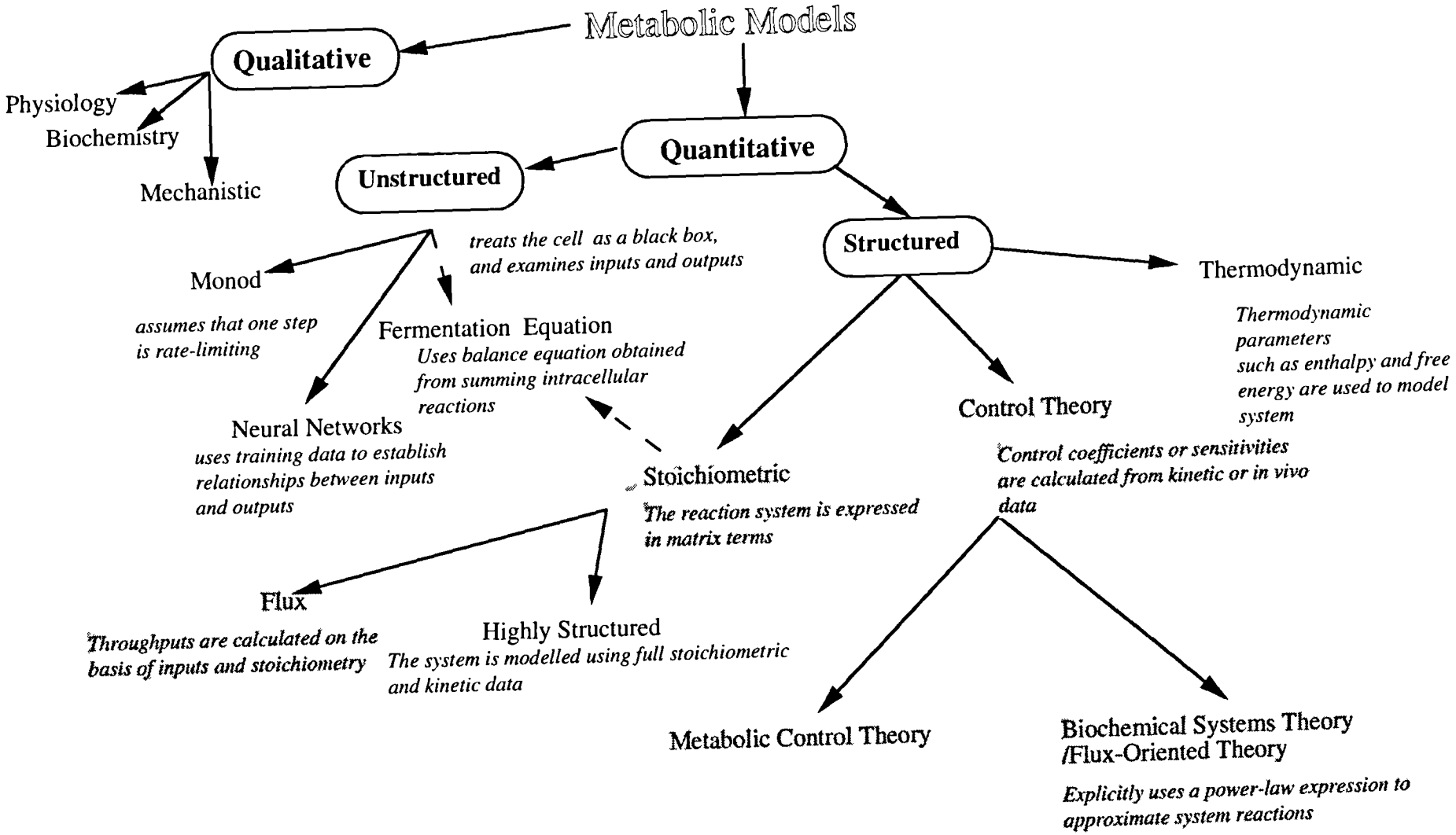


Figure 1.1: Schematic representation of different approaches to modelling metabolic systems. The techniques applied to the TOL pathway are highlighted.

1.1 Metabolic Pathway Engineering Without the Use of Control Theories

A prerequisite for pathway engineering is a good knowledge of the system being altered. It is necessary to know what changes should be made and the effects of these on the overall metabolism of the cell. Below, the pathway engineering of several different systems is described, including those whose mechanisms are well-established and one where pathway engineering work is just beginning. Following this, the way in which modelling techniques can be used to provide insight into and verification of cell mechanisms and direct cloning experiments is discussed.

1.1.1 A Knowledge-Based Approach:

1.1.1.1 Optimization of the Flux in the Phenylalanine Pathway:

The production of the amino acid phenylalanine has been extensively studied especially due to its commercial success, but since much of this has been in the industrial sector there has been relatively little published. However, one study by Backman *et al* (1990) is a very good example of an experience-based approach to metabolic pathway engineering.

A diagram of the phenylalanine biosynthetic pathway showing the feedback repression mechanisms is given in Figure 1.2. Backman *et al* (1990) took the approach of cloning all of the genes involved in the pathway except for the transaminases. Each of the genes was then engineered for increased expression and the effect of such increased expression was measured. This enabled them to determine which genes were important in the rate of phenylalanine synthesis. As expected not all of the cloned genes increased phenylalanine production rate and one gene even had a negative effect, decreasing phenylalanine titres by 20%. This negative effect on flux or "wrong way behaviour" is not unusual in biochemical systems and occurs as a result of such complex structures as substrate cycling and metabolite conservation. (A discussion of the control analysis of such systems is given in Section 1.2)

Having done this they then concentrated on removing the effects of regulation by

phenylalanine from the pathway. There are two ways that this can be done: either by changing the regulation of enzyme activity (e.g.by site-directed mutagenesis) or by altering the regulation of gene expression. Three points at which regulation occurred were identified as shown in Figure 1.2. The first step in the biosynthesis is catalyzed by three isoenzymes encoded by the genes *aro*F, *aro*G, *aro*H and regulation of this step occurs both at the level of control of transcription of these genes and at the level of the enzyme activity of DAHP synthetase (by feedback inhibition). The second point of control is in the formation of shikimate kinase which is encoded by *aro*L and the third point of control occurs at the conversion of chorismate to phenylalanine where the enzyme chorismate mutase-preprenate dehydratase (CMPD), encoded by the gene *phe*A is regulated both at the level of enzyme formation and enzyme activity.

Both of the approaches mentioned above (those of changing both the enzyme activity and gene expression) were used to maximize the flux through the pathway by removing the mechanisms of regulation, as follows:

(i) by generating mutations in *tyr*R repressor protein gene:

Because several of the genes in the pathway are regulated by the repressor protein encoded by *tyr*R, the mutations generated here removed repression on the formation of the enzymes encoded by several genes, including *aro*F, *aro*G, *aro*L and genes involved in the transport and transamination of phenylalanine.

(ii) by replacing the natural promoter for *phe*A to avoid transcriptional regulation:

The regulation of the *phe*A gene is sufficiently complex that it was decided to replace the entire promoter rather than deal with each of the forms of regulation individually. Several promoters were tried and the final choice was one that gave the highest expression level.

(iii) by deletion of a residue from CMPD so that it was no longer subject to feedback regulation:

In the feedback repression of CMPD by phenylalanine, a tryptophan residue is essential so that its substitution or deletion results in the inactivation of the repression mechanism. This is a good example of how protein engineering by site-directed mutagenesis can by used to alter the control of metabolism.

(iv) by isolating feedback-inhibition-insensitive mutants in the gene for DAHP synthase by means of resistance to toxic amino acid analogues.

By using these techniques the production of phenylalanine was increased to 50 g/l. Other efforts at maximizing the productivity of the phenylalanine pathway have resulted in titres of up to 80 g/l and at this level the phenylalanine crystallizes out of solution. This is one of the most successful examples (in the public domain, at least) of how genetic engineering techniques, along with a thorough understanding of the molecular biology and regulation of a system, can be used to engineer a metabolic pathway and it gives an indication of the potential for over-production in biological systems.

1.1.1.2 Where To Start On A Complex System: The Penicillin Pathway:

Having discussed above one pathway which has been successfully engineered and before going on to examine a (relatively simple) pathway on which there is enough information to begin a pathway engineering study, it is useful to consider how we should begin to tackle the pathway engineering of an industrially very important, but extremely complex system such as the penicillin biosynthetic pathway.

As with any project it is necessary to define the eventual goal, even if the long-term aim seems many years from being achieved. In pathway engineering terms, it will usually be one of the following:

(i) increased productivity

e.g by maximizing flux through the pathway

- by changing the fermentation conditions
- by changing the host

(ii) the development and production of novel antibiotics

- by installing new pathways
- by altering the pathways present
- by changing the host

Examining the first heading, the term "productivity" is taken to include the titre of penicillin in the broth, the efficiency of conversion of carbon source to product and cost-effectiveness with respect to fermentation conditions and downstream processing. A general scheme for the biosynthesis of penicillins and cephalosporins in filamentous

fungi and *Streptomyces* is shown in Figure 1.3. The precursors for this pathway are L-aminoadipic acid, L-cysteine and L-valine. This pathway has been studied in detail at the enzymatic level so that there are data about how the individual reactions occur *in vitro*. The challenge is to relate this to the action of the pathway *in vivo*. To do this requires detailed knowledge of the control of the system at the transcriptional, translational and enzymatic level.

The first hurdle which must be overcome is the complete genetic characterization of the molecular biology of antibiotic synthesis. In all systems so far studied, antibiotic genes have occurred in clusters. Indeed for most of the systems where all of the biosynthetic genes have been accounted for, or where heterologous expression of antibiotic production has been achieved, all of the antibiotic genes have been found in one cluster often along with a gene encoding for resistance. Examples include actinorhodin, erythromycin, cephamycin C, tetracenomycin C and oxytetracycline (Hunter and Baumberg 1989). An exception to this was reported by Skatrud and Queener (1989) who found that the IPNS and DAOCS genes of *C.acrimonium* reside on different chromosomes.

To date the genes for three of the enzymes in the pathway have been individually cloned: those for ACVS (Smith *et al* 1990a), IPNS (Samson et al 1985) and REXH (Samson *et al* 1987). In addition a cluster containing the genes encoding ACVS, IPNS and ACT activity from *P.chrysogenum* has been cloned and expressed in *Neurospora crassa* and *Aspergillis niger* (Smith *et al* 1990b).

The occurrence of these genes in clusters may give the impression that their transcription and its control should be quite straightforward but this would be misleading. The mechanisms of transcription and their controls are quite complex and it is not even clear what controls the initiation of secondary metabolite production. Also most gene clusters for antibiotic production are located on the chromosome rather than on plasmids which makes cloning more difficult, except where stable transformants are available. (The one well-documented exception to this is methylenomycin A (Hunter and Baumberg 1989).)

Antibiotic production begins after the initial rapid growth phase and correlates with the appearance of the enzymes of their biosynthetic pathways (the exception to this being IPNS which is present from the beginning of the fermentation). This is an important

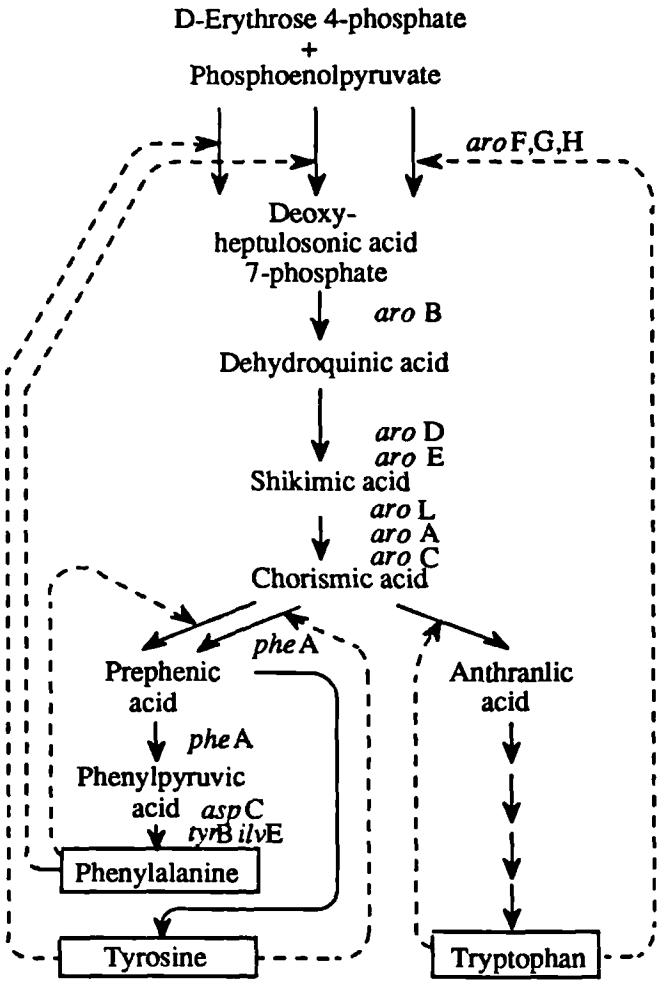


Figure 1.2: Phenylalanine biosynthetic pathway, showing points at which feedback repression occurs.

point as it implies that it is the repression of transcription of the genes encoding these enzymes which occurs in the initial growth stages rather than repression of the enzyme activity. The details of this repression have not been worked out but there is a large amount of empirical information on the effect of fermentation conditions (Bushell 1989). In general, secondary metabolite production is enhanced by the use of carbon and nitrogen sources which permit only slow growth. It is repressed by high phosphate concentrations and high growth temperature. In nature it seems likely that secondary metabolite production was connected with the formation of ariel mycelium and this has lead to hypotheses that they play a role in cellular differentiation. One of these hypotheses implicates guanine nucleotides in the control of antibiotic production. It has been shown (Ochi 1984, 1987) for *Bacillus subtilis* and *Streptomyces griseus* that sporulation is controlled by levels of one or more of GMP, GDP or GTP and ppGpp controls the level of antibiotic production. Care must be taken when trying to extend this result however. Factors such as the selection marker on the cloning vector used can influence the control mechanism, for example, thiostrepton which prevents ppGpp formation.

The transcription patterns of antibiotic synthesis pathway clusters are very complex. They are composed of several transcriptional units: 3 in the case of methylenomycin and at least 4 and 7 for actinorhodin and streptomycin respectively. Multiple sites of transcription initiation are present and, in the case of antibiotic resistance genes, multiple holoenzymes can be used to transcribe the same gene. One explanation for this is that it gives a form of temporal control so that production occurs only at some stages in the cells growth and not at others. Also, divergent transcription has been observed in many antibiotic resistance gene clusters.

Catabolite repression obviously plays an important part in the regulation of antibiotic synthesis genes. Secondary metabolism is repressed by carbon, nitrogen and phosphate and one model for this mechanism involves a low molecular weight effector whose concentration reflects the physiological state of the cell which modulates the activity of a regulatory protein. One candidate for this in the streptomycetes is A-factor which activates streptomycin production and resistance (and also sporulation) in *S.griseus* and *S.bikiniensis*. However it has been shown that A-factor formation is itself regulated by the product of two genes *afs*B and *afs*C in a scheme of the type shown in Figure 1.4.

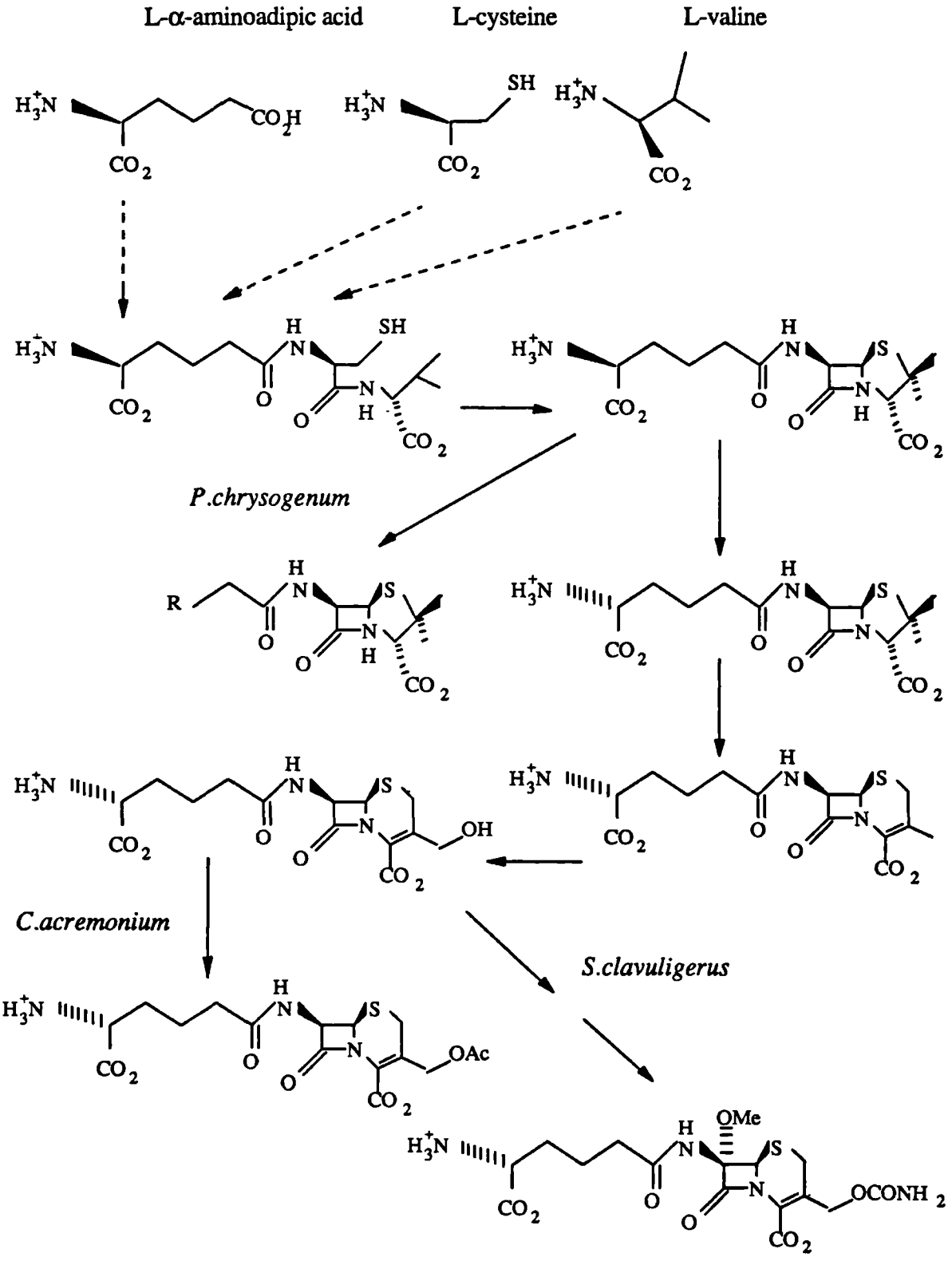


Figure 1.3: Biosynthesis of penicillins and cephalosoprins in filamentous fungi from the precursors L-α-aminoadipic acid, L-cysteine and L-valine.

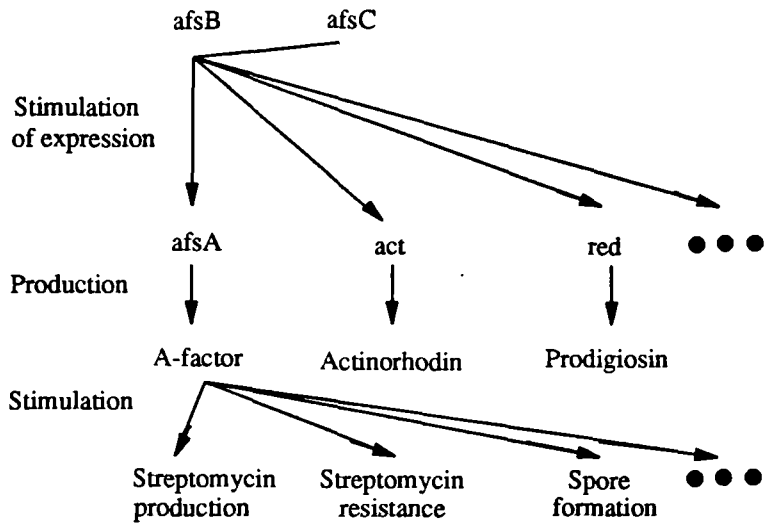


Figure 1.4 Regulation of A-factor formation

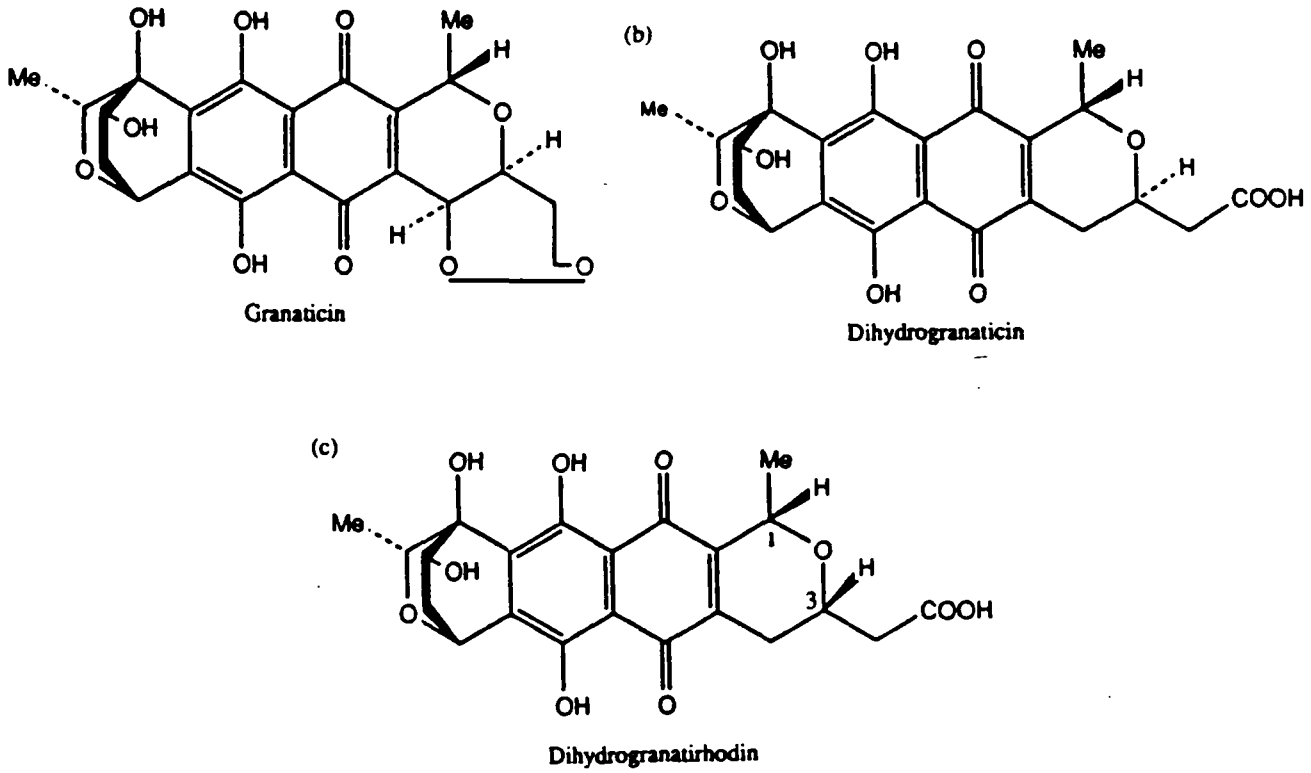


Figure 1.5: Structure of the novelantibiotic, dihydrogranatirhodin

There may be also a spatial factor as in *Penicillium chrysogenum* where it has been shown that the biosynthetic machinery for penicillin G is located in Golgi-type vesicles whose formation corresponds to the time of synthetic activity (Kleinkauf and von Dohren 1990).

Faced with the complexity of the above problem where should one begin with the engineering of this pathway in order to, say, increase the flux to product? The procedure might be as follows:

(i) Increase flux to precursors:

At the moment a standard antibiotic fermentation has a carbon source to product conversion factor of only about 8%. It is obvious that if we wish to increase this figure we must set about redirecting the carbon flow towards the pathways for antibiotic synthesis. To do this we should begin with the precursors.

It is not known in detail what effect the levels of precursors of the pathway have on its regulation but it has been shown that feeding a precursor to an antibiotic fermentation can increase yield. This has the disadvantage that it must be transported into the cell and the added complication that there seems to be some interaction between transport mechanisms and antibiotic production. In the case of β-lactam antibiotics the precursors are L-α-aminoadipic acid, L-cysteine and L-valine. With the increasing information available on the primary metabolism of antibiotic-producing organisms it should soon be feasible to engineer a strain for increased precursor production. Alternatively, traditional mutagenesis and screening techniques could be used to screen for over-producers of the precursors.

(ii) Identify the rate-limiting enzymes:

If we can ensure that the precursors for the pathway are present in excess either by feeding, by engineering the pathway or by screening for mutants which overproduce the precursors, the next problem is to identify the rate-limiting enzymes in the pathway and to increase the activity of these enzymes. There are several ways of doing this. One is to do a flux analysis of the pathway (see Section 1.1.3.1.1) and identify the bottlenecks. The rule of thumb that it is often the enzymes directly after these bottlenecks which are rate-controlling can then be used (Bailey and Ollis, 1986). Because of the complexity of the control of these systems it is possible that a more detailed control analysis may

be required and these methods are described in Section 1.2. A more *ad hoc* method is to increase the expression of each of the enzymes in the pathway as was done for the phenylalanine pathway in Section 1.1.1.1. This can be equally effective but can be wasteful in that (a) not all of the enzymes in the pathway may have been cloned and (b) a lot of time and effort could be put into the cloning and expression of an enzyme that might have little or no effect on the rate through the pathway.

To complicate the issue it now seems as if the enzymes which are rate-limiting in the antibiotic production pathways vary from organism to organism. For example an increase in IPNS causes an increase in the production of penicillin in *P.chrysogenum* (Earl 1990) whereas in *C.acrimonium* an increase in IPNS shows no effect but over-expressing REXH causes a 30-80% increase in yield (Pratt 1989). Without a mechanistic model for the control of the pathway it is difficult to explain these differences so until our understanding of the system improves we can only be aware of the dangers of making generalizations.

(iii) Minimize catabolite repression:

We know that antibiotic production is repressed by high levels of carbon, nitrogen and phosphate. At the moment this effect is minimized by the use of low-quality feedstocks such as corn-steep liquor and by careful feeding regimens. Because of the poor understanding of the mechanisms of repression this will be a difficult problem to overcome. One approach is to isolate repression-insensitive mutants (Chang *et al* 1990). These often perform poorly in other respects and so are not useful industrially but are useful to studies of repression. A study of the fluxes through the pathway and the control exerted by the enzymes at different levels of repression could indicate which enzymes were influenced by what types of catabolite repression and therefore suggest solutions.

More work is required to verify the regulation model described in Figure 1.4 but if it is true then work could begin on increasing the expression of the proteins *afs*B, *afs*C and *afs*A.

(iv) Control of initiation of antibiotic production:

Industrially it is important to be able to control and predict the stage in the fermentation

at which antibiotic production occurs. Unfortunately there is little knowledge of the details of the mechanism of initiation of transcription. As mentioned above ppGpp has been implicated in this mechanism but exactly what role it plays is not clear. Because of the connections between cellular differentiation and onset of secondary metabolite production it is likely that developments in these two areas will proceed concurrently. Structural models which seek to improve our understanding of such mechanisms by simulations could prove invaluable in these areas and an example which models cellular differentiation in *B.subtilis* is discussed in Section 1.1.3.1.3.

The engineering of antibiotic biosynthetic pathways for the production of novel antibiotics requires a slightly different approach. Here we are trying to change the rules which the organism has been programmed to follow and there are several methods which can be applied.

(i) Use of precursor analogues:

The use of precursor analogues containing functional groups which we wish to be present in the final product is possible because of the relaxed substrate specificity of some of the enzymes of the pathways. For example there are many analogues of L-α-aminoadipic acid which can be transformed by IPNS (Pratt 1989). REXH, however, is much less promiscuous in its substrate specificity as are several of the other enzymes in the pathway. Because of the developments in modern genetic methods we can now alter the activity of the enzyme itself using techniques from protein engineering such as site-directed mutagenesis [Baldwin et al, 1989].

(ii) Introduction of new DNA to create new pathways:

There are several ways to introduce non-native DNA into a microorganism. Recombination between species has been attempted using protoplast fusion but this requires almost totally homogenous DNA for recombination to proceed efficiently. The introduction of heterologous DNA on autonomously replicating plasmids seems to have more potential. Using this method the novel antibiotic dihydrogranatirhodin (shown in Figure 1.5) was produced following the introduction of the actinorhodin gene cluster into *S.violaceoruber* which normally makes granaticin or dihydrogranaticin. In a similar way systems have been developed which can alter the backbone structure of an

antibiotic (for an example in *S.galilaeus* see Hunter and Baumberg 1989) and the potential in this area seems great.

1.1.2 Creation of New Pathways:

1.1.2.1 Production of Indigo:

A new route to the bacterial production of the dye indigo was proposed by Mermod *et al* (1986) who showed that cells of *E.coli* K-12 containing a cloned fragment of *Pseudomonas putida* TOL fragment pWWO produced indigo.

Indigo is produced by *Pseudomonas* sp. bacteria by the oxidation of indole to indoxyl by molecular oxygen. It was shown that indigo is formed in *E.coli* cells containing cloned genes encoding a naphthalene dioxygenase enzyme of *P.putida*. The proposed pathway, shown in Figure 1.6 involves the formation of indole from tryptophan,by tryptophanase, followed by the formation of *cis*-indole-2,3-dihydrodiol from indole by naphthalene dioxygenase. Spontaneous dehydration of *cis*-indole-2,3-dihydrodiol results in the formation of indoxyl which in turn oxidises spontaneously to form indigo. This reaction mechanism suggested to Mermod *et al* that a *Pseudomonas* TOL plasmid-specified hydroxylase, xylene oxidase, would form the same product.

This is rather typical of many pathway engineering efforts as it is not quite clear if this discovery was entirely by chance (the production of indigo having a very obvious phenotype) with the biochemical details applied afterwards to explain the phenomenon. Obviously not all novel pathways can be so easily identified and a systematic methodology is essential if the search for novel pathways or products is to be optimized.

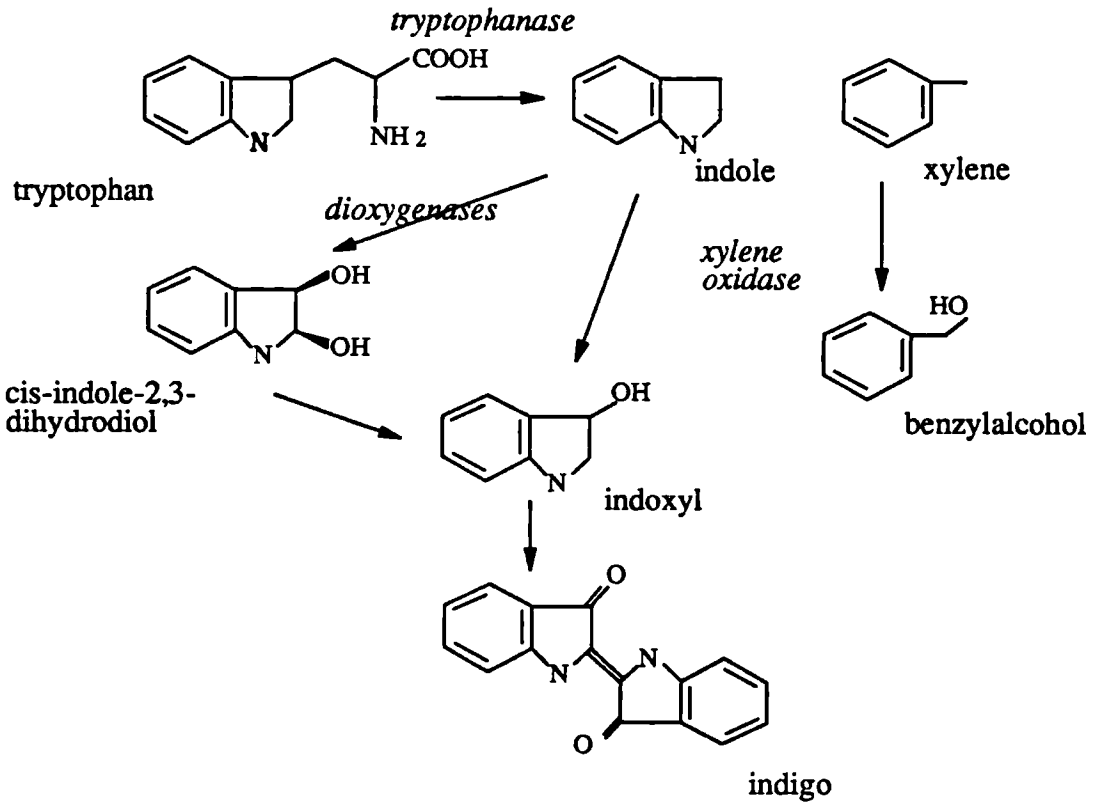


Figure 1.6: Formation of indigo from tryptophan in *E.coli* by the action of a non-native oxidase.

1.1.2.2 Helping the cell to adapt to its processing environment:

This involves changing the cell so that it can cope better with the conditions under which we wish to grow it. This is normally done by selecting for mutants with the required properties which is essentially letting the microorganisms genetically engineer themselves and then choosing those that have been most successful, but this method has its limitations, depending on the availability of an efficient screening strategy.

An alternative method is to confer a desirable phenotype upon the strain using genetic engineering methods. An example of this is the use of a haem-containing oxygen-binding protein synthesized by a bacterium of the genus *Viteroscilla* to enable *E.coli* to utilize oxygen more efficiently under microaerobic conditions. This work was carried out by Khosla and Bailey (1988) and involved cloning the 2.2-kb *Hin*dIII fragment coding the protein into *E.coli* JM101 via a pUC plasmid. The cells were shown to express the protein by 2-D gel electrophoresis. Cells containing the gene grew to higher cell densities than either plasmid-carrying or plasmid-free controls.

The advantage of this approach is that it frequently works. The main disadvantage is that it is difficult to tell in advance whether or not the chosen cloning strategy will be successful. It is very much a trial and error approach. If the system is too complex to conceptualize but the correct cloning choices are made then that is termed intuition if the worker is experienced, or luck if he is not. Another drawback is that while specific experiments are admittedly the only way to prove or disprove a hypothesis they fail to yield general insights into the structure of metabolic pathways. Also it is difficult to know where to begin to study novel or complex systems with this approach.

For this reason we need a framework to help us to visualise metabolic systems and predict likely outcomes. These frameworks are called models.

1.1.3 Metabolic Models and Expert Systems

1.1.3.1 Models - From Stoichiometric to Structured

A model is a formal representation of a material system. It can be as simple as a diagram or as complex as a system of partial differential equations but its object is to predict, simulate or explain experimental observations of the system.

The use of models allows us to simulate metabolic systems. These simulations can provide insight into the behaviour of such systems and a basis for control analysis. By that we mean an analysis of what regulates the system in terms of the rate-controlling steps and enzymes. It provides a framework upon which to fit available experimental data as well as a means of predicting quantities in a system. Models are necessary to provide a basis for experimental design - that is, determining which experiments are necessary and what variables must be measured.

Before we begin modelling a system we must decide what parts of the system we wish to include and to what degree of complexity.

Choosing which parts of a system to include will require intelligent simplification and reduction of the complex network of reactions present in any cell. The metabolic pathways which must be considered in the model will vary with the fermentation conditions (for example, the carbon source will determine which catabolic pathways are switched on) and with the genotype (whether or not any new pathways are being expressed in the system, for example). Deciding on the degree of complexity of the model will require careful consideration of the objectives. To carry out a control analysis of a system will require the steady state metabolite concentrations and fluxes as initial data, whereas an analysis of the carbon flow through the system would require knowledge of the fluxes only.

The descriptions given below of the different areas of modelling are necessarily brief: the intention is to point out the uses and applications of the different approaches rather than give a detailed description of how to employ them. The relationship between the different types of models is shown in Figure 1.1.

1.1.3.1.1 Flux Modelling:

The minimum knowledge which we can expect to have of a system is some of its metabolic pathways. These are well established for most microorganisms, especially for those commonly used such as *E.coli* or *Bacillus* species. Given just this framework, along with measurements of carbon consumption, CO_2 evolution, excreted products and biomass during a fermentation we can form a picture of the carbon flow through the system. This method has been used by several authors for purposes varying from the identification of a correct model (Aiba and Matsuoka 1979) to studies of the carbon flow during growth on different substrates (Holms 1986,1987, El-Mansi and Holms 1989, Nimmo *et al* 1989).

Because this method is based upon measuring the inputs and outputs and performing a mass balance across the system, it is perhaps not surprising that most of its applications concern the central carbon metabolism. Aiba and Matsuoka (1979) used it in their analysis of citrate production by *Candida lipolytica* from glucose, where they investigated the validity of three different models. Their base metabolic model is shown in Figure 1.7 and of the three models which were considered, the first coordinated the pyruvate carboxylation with the TCA cycle and disregarded the glyoxylate cycle, the second overlooked pyruvate carboxylation and the third overlooked 2-oxogluterate dehydrogenation.

Measurements were made of the concentrations of cell mass, protein and carbohydrate, citrate, isocitrate, α-ketogluterate and glucose in NH_4-limited chemostat culture. $C0_2$ evolution and NH_4^+ uptake were also monitored. Writing out the carbon balance equations for each metabolite yields the 11 branch point constraints in Table 1.1 in terms of the carbon fluxes $v_1,\ldots,v_{12}$ shown in Figure 1.7.

The number of unknown carbon fluxes can be reduced from 12 to 7; v_{12}, v_1, v_4, v_9 and v_5 can be deduced from experimental data. The number of independent relations can be reduced to 6 (this concept of independent reactions is very important in the solution of metabolic models and is discussed below in Sections 1.1.3.1.2) so that we must eliminate one more unknown before attempting to solve the system. This is the basis of the three models above. In model 1 we set $v_7 = 0$, in model 2 $v_3 = 0$ and in model 3 $v_8 = 0$.

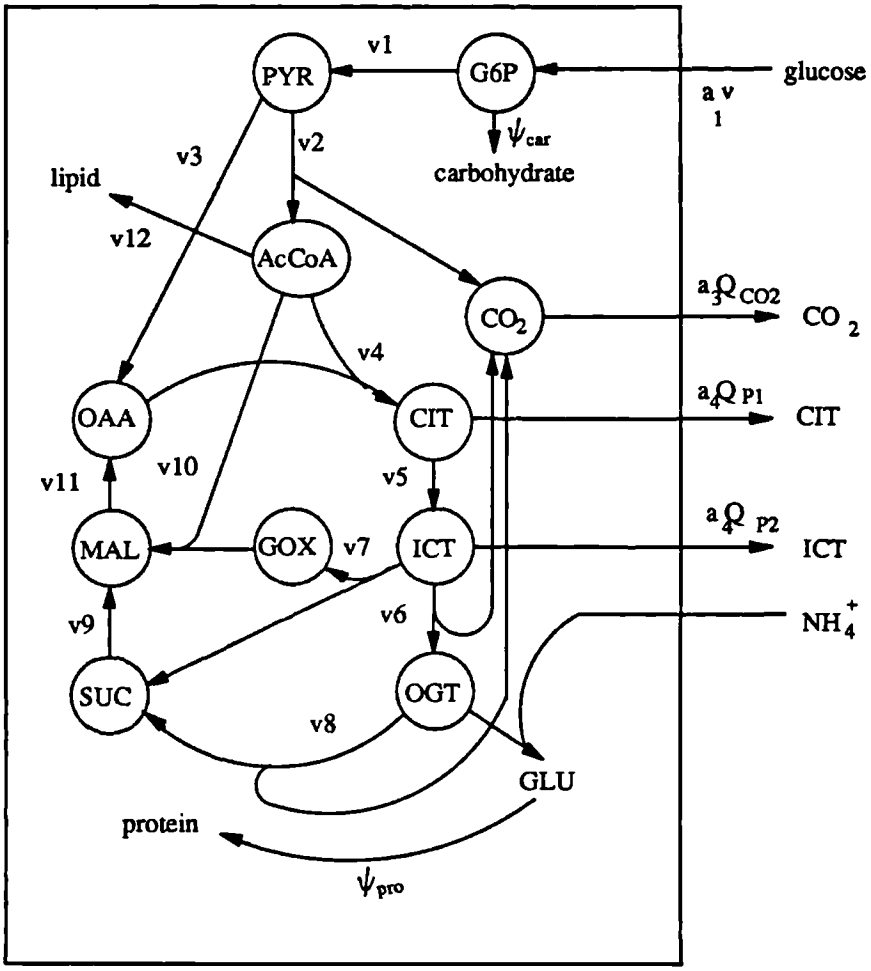


Figure 1.7: Metabolic model showing citrate production from glucose

Table 1.1: Carbon balance equations resulting from the metabolic model above.

Metabolite Pool		Carbon Balance Equation
G6P	$\alpha_1\nu$	$= v_1 + \alpha_7\psi_{car}\mu$
PYR	v_1	$= v_2 + v_3$
AcCoA	$2/3\ v_2$	$= 1/3\ v_4 + 1/2\ v_{10} + v_{12}$
CIT	v_4	$= v_5 + \alpha_6 Q_{P1}$
ICT	$5/6\ v_6$	$= v_8 + \alpha_6\psi_{pro}\mu$
OGT	$2/3\ v_7 + 4/5\ v_8$	$= v_9$
GOX	$1/3\ v_7$	$= 1/2\ v_{10}$
MAL	$v_9 + v_{10}$	$= v_{11}$
OAA	$4/3\ v_3 + v_{11}$	$= 2/3\ v_4$
CO_2	$1/3\ v_2 + 1/6\ v_6 + 1/5\ v_8$	$= 1/3\ v_3 + \alpha_3 Q_{CO2}$

The carbon fluxes can now be calculated from the experimental data and the models compared. It was found that in models 2 and 3 negative fluxes occurred. This contradicted what was known from the free-energy changes of these reactions which predicted forward reaction directions. These results suggested that models 2 and 3 be discarded. This was confirmed by measuring the enzyme activities of the pathways.

The above experimental process is a very good example of how the different pieces of information about a system must be brought together in order to validate or invalidate a model. It also illustrates how it is necessary to check the results of a simulation against our knowledge of the system and how it behaves experimentally.

The same principles of mass balance and flux calculation can be use to investigate, in qualitative terms, the control properties of a pathway. That is, what are the regulatory mechanisms used by the cells and where in the pathway are they used. One common regulatory mechanism is end-product repression either of the expression or the activity of the enzyme. One general rule is that the junctions in the central pathways at which flux is divided among several outputs are likely to be controlled by one or more of the regulatory mechanisms available to the cell. We can calculate the flux through each enzyme in a pathway by multiplying the throughput of each step (defined as mmol of carbon transformed to give one gram of dry weight) by the growth rate (μ) (Holms 1986). The throughputs are obtained from experimental data on the carbon inputs and outputs to biosynthesis, CO_2 and the excreted products.We can thus determine where such junctions occur.

This approach has been used to investigate the central metabolic pathways of *E.coli*, in particular the control of carbon flux to acetate excretion (Holms 1986, El-Mansi and Holms 1989) and the control of flux through the citric acid cycle and the glyoxylate bypass (Holms 1986, 1987).

In the study of the excretion of carbon source to acetate, it was determined from the flux model that the key junction was at pyruvate. The fluxes through this junction will vary with the carbon source as the organism attempts to bridge the gap between the phosphorylated and non-phosphorylated parts of the central pathways. This provides an indication of how we should proceed if we wish to engineer this pathway for greater efficiency to biomass. There are two approaches which can be taken: the reduction of the flux into the pyruvate junction by controlled feeding of glucose, the running of a

glucose-limited culture (El-Mansi and Holms 1989) or the reduction of the flux to acetate by altering the activity of the enzymes in that pathway (Diaz-Ricci and Regan, 1991).

The growth of *E.coli* on acetate provides an opportunity to investigate the control of flux through the citric acid cycle and the glyoxylate bypass. The function of the glyoxylate bypass is to provide carboxylic acids and phosphorylated intermediates using only Acetyl-CoA as a source. By performing a flux analysis, it can be shown that the control of flux through this bypass will most probably lie with the enzymes isocitrate lyase (ICL) and isocitrate dehydrogenase (ICDH) since the synthesis of all of the biosynthetic precursors depends on the flux through ICL. This control mechanism has been shown to depend on reversible inactivation of ICDH by a bifunctional kinase/phosphatase. This kinase/phosphatase responds to two types of effectors - the intermediates which are generated by flux through ICL and ADP, AMP and $NADP^+$. This has the effect of adjusting the flux through the citric acid cycle so that the rate of production of ATP and NADP is equal to the demands of biosynthesis. (Holms 1986, 1989)

This type of approach is essential when beginning to study any system, either from the view of control analysis or pathway engineering but it is limited and in many cases may provide no more than a starting-point from which to apply other methods of analysis. It provides no data how control is distributed over the enzymes in a pathway before and after a critical junction, for example. It is useful as a means of devising an optimum feeding program so that the fluxes to the desired product are maximized (Holms *et al* 1991). Genetic engineering of these pathways, however, could increase these fluxes by orders of magnitude and for this we need to identify more precisely the enzymes with which the control of a pathway lies. We will see how this may be done in Section 1.2.

1.1.3.1.2 Stoichiometric and Pathway Methods:

Much can be learnt about a system by examining the stoichiometric constraints across it. These constraints include carbon, nitrogen and available electron balances. While these types of models have been primarily used in the monitoring and control of

bioprocesses they can also be used to validate metabolic pathway models and to model any parts of a system which we wish to consider as a black box. These types of stoichiometric methods are also very instructive in illustrating the difference between independent and dependent reactions, a concept which is important in all systems of reactions. They can also be used to determine how many flux measurements are required in order to solve for a system.

A "beginners guide" to this kind of steady-state simulation has been written by Hofmeyr (1986). This work brings together the conclusions of several other authors but is convenient because of its clarity.

The first step is to write down the stoichiometric matrix. Writing the steady state in terms of the stoichiometric matrix and its rate expressions is the basis of the solution of many of the structured models discussed below. The stoichiometric matrix is a way of describing the system reactions numerically. It is a matrix of m rows and r columns where m is the number of metabolites and r is the number of reactions. The entries a_{ij} in the matrix are called the stoichiometric coefficients and represent the number of molecules of metabolite s_i participating in the stoichiometric equation of the reaction j i.e.

$$\boldsymbol{A}^T\boldsymbol{x} = 0 \qquad (3)$$

where x are the metabolites in the reaction system. Writing the balance equations for each metabolite

$$v_i = \sum_{j=1}^{r} a_{ij} v_j \qquad i = 1, \ldots, m \qquad (4)$$

where X is the concentration of the metabolite x and v_j is the rate of reaction j. At steady state the metabolite concentrations are time-invariant so that we have the matrix equation

$$\boldsymbol{A}\boldsymbol{v} = 0 \qquad (5)$$

This is a non-linear problem since the rate vector v is a function of substrate, product, effector and enzyme concentrations as well as kinetic constants. This is the system which we must attempt to solve. We begin by reducing the number of variables. This is normally possible because of the presence of substrates which are in excess or which are rapidly removed, either by reaction or transport processes. In addition conserved metabolites (or rather their sum) such as NAD^+ and NADH can be expressed as a constant. The rows of A corresponding to these constant metabolites can now be deleted to give a reduced matrix A' of dimensions m'x r.

If v is a linear vector, a steady state is possible only if the rows of this matrix are linearly independent, that is, no row may be a linear combination of the other rows. The number of linearly independent rows is given by the rank r' of the matrix. (The rank is given by the number of rows containing non-zero elements when the matrix has been reduced to its echelon form by Gaussian elimination.)

If m'=r' then the system is determined and a steady state can be computed. It should be noted that this steady state is not necessarily stable or unique.

If moiety conservation occurs then m' < r' and the system will be undetermined. In this there should be m'- r' conservation equations with which we can replace the appropriate balance equations in the matrix. In this way we reduce the matrix so that m'=r'. and the steady state solution can be reached.

This concept of whether the system is ill-determined recurs frequently (see Sections 1.1.3.1.3) in the study of metabolic systems and it is important to have a clear idea of what is variable in a system and what is fixed so the problem does not become under or over determined.

If we elaborate slightly upon the above structure its uses expand. If the m compounds are composed of q atomic species then their molecular formulae can be described by the equation

$$\boldsymbol{x} = \boldsymbol{B}\boldsymbol{b} \qquad (6)$$

where **B** is an m x q matrix of atomic coefficients and **b** is a q x 1 vector of atomic species

We can obtain n_i, the number of moles of species i produced in the overall reaction from

$$n_i = \sum_{j=1}^{r} a_{ji}\xi_j \qquad i = 1,\ldots,m \qquad (7)$$

or in matrix form

$$\mathbf{n} = \mathbf{A}^T\boldsymbol{\xi} \qquad (8)$$

where **n** is the m x 1 column vector of flows and $\boldsymbol{\xi}$ the r x 1 column vector of extents of reactions.

The maximum number of independent reactions involving m compounds composed of the q atomic species is then given by the Gibbs rule of stoichiometry (Niranjan and San 1989, Tsai and Lee 1988) as

$$r_{max} = m - q + g \qquad (9)$$

where g is the number of relations observed between the columns of **B**. In biological systems there are frequently only four atomic species of any importance: carbon, oxygen, nitrogen and hydrogen so that q = 4 and g = 2 and the maximum number of independent reactions is two less than the number of compounds in the system. It can be shown that, if there are l external fluxes, then

$$r + l = m - q + g \qquad (10)$$

is a necessary and sufficient condition to solve the system uniquely for any values of the m - q - l experimentally determined values and this proof is given in Appendix C. Using this framework, Niranjan and San (1989) examined the conditions to delineate among cellular mechanisms based solely on extracellular measurements. (The aim of

this work is similar to that of Aiba and Matsuoka (1979) described in the previous section.) They arrived at the conclusion that when $r = r_{max}$ there are no constraints among the stoichiometric coefficients in the overall reactions apart from those imposed by the elemental balances and for this case it is not possible to uniquely determine the intercellular mechanism given only extracellular measurements. They also made the point that no information on any mismatches between the model pathways and the actual mechanisms could be obtained from extracellular measurements.

As an alternative to the elemental balances described by Eqn. 6 we could write a conservation equation over the carbon and reductance degree. This is an unstructured type of model as it treats the cell as a black box, not using any of the biochemical data on the system. This balance of the number of available hydrogens (or available electrons) is obtained from the oxidation of the compound to CO_2 and NH_3 with water (Papoutsakis 1984, Papoutsakis *et al* 1985a,b).

For example, the stoichiometric balance equation for C, H, O and N in an anaerobic fermentation can be written in the form

$$CH_mO_l + aNH_3 = y_cCH_pO_nN_q + zCH_rO_sN_t + bH_2O + cCO_2$$

where CH_mO_l, $CH_pO_nN_q$ and $CH_rO_sN_t$ denote the elemental compositions of the organic substrate(s), microbial biomass and extracellular products respectively. p, n and q are found from the elemental analysis of the biomass and m, l, r, s and t are known from the molecular formulae of the substrate and products.

In the available electron balance each H in excess contributes -1 to the electron value of the compound: a lack of H contributes +1 and each excess N is worth +3. The electron values of NH_3 and H_2O are zero and the electron balance for the above expression yields (after multiplication by 2)

$$2l - m = y_c(2n + 3q - p) + z(2s + 3t - r) + 4c \qquad (11)$$

Writing the balance for the number of available hydrogens obtained from the oxidation of the compound to CO_2 and NH_3 with water we have

$$4 + m - 2l = y_c(4 + p - 2n - 3q) + z(4 + r - 2s - 3t) \quad (12)$$

Writing this balance in degrees of reductance of the compound (γ), defined as the number of equivalents of available electrons per atom of carbon in the compound

$$\gamma_s = y_c\gamma_b + z\gamma_p \quad (13)$$

Tsai and Lee (1988) have shown that this method can yield extra information and reduce the size of the system to be solved when $r < r_{max}$ i.e. when the number of independent reactions is not at its maximum. Otherwise (at $r = r_{max}$)the usefulness of this method is limited and it should be applied with caution.

We can apply statistical analysis to this framework to avoid the possibility of false conclusions due to errors in the experimental data. Tsai and Lee (1988) have shown that the maximum likelihood solution is given by

$$\boldsymbol{x}^* = (\tilde{\boldsymbol{A}}_{ext} \cdot \boldsymbol{F}^{-1} \cdot \tilde{\boldsymbol{A}}_{ext}^T) \cdot \tilde{\boldsymbol{A}}_{ext} \cdot \boldsymbol{F}^{-1} \cdot \boldsymbol{n}^+ \quad (14)$$

where $\mathbf{x}^*$ is the maximum likelihood estimate of $\mathbf{x}$ and $\mathbf{n}^+$ is the measured value of $\mathbf{n}$. $\mathbf{F}$ is a covariance matrix which, if the measurement errors are uncorrelated and of the same magnitude, can be expressed as

$$\mathbf{F} = c\,\mathbf{I} \quad (15)$$

where c is a constant representing the error.

1.1.3.1.3 Structured Models:

What differentiates highly structured models from those previously described is that the metabolic pathway is described in as much detail as possible and each reaction in the pathway is modelled by a kinetic expression which should contain details of any

interactions which may occur involving other metabolites. Such models can be expanded as is necessary, for example, to include details on transcription and translation (Peretti and Bailey 1986) or on initiation of sporulation (Jeong and Ataai 1990). It is this aim of describing in detail the complex interactions of the cellular mechanism which sets this branch of modelling apart.

To set up a structured model it is first necessary to write down the metabolic pathways, in the same way as for the flux model shown in Figure 1.7. The conservation equation for each metabolite is expressed as an ordinary non-linear differential equation with each reaction and interaction described by a kinetic expression. This system must then be solved using numerical solution techniques. For simple reaction schemes, such as those in some chemical engineering processes this is a straightforward method which yields easily-interpreted results. However microbiological systems are sufficiently complex that there are some difficulties which may arise.

The first of these is in deciding which elements of a metabolic system to include and to what degree of complexity. This will mostly be determined by the hypothesis which one is trying to prove or the control system which one is trying to elucidate. Typically, in modelling whole cells, one would include the pathways of central carbon metabolism, pathways to amino acid and protein synthesis and details of purine nucleotide metabolism.

The second difficulty arises in parameter estimation. These types of models normally require a large number of variables, most of which will have to be taken from *in vitro* data. A degree of judgement is required to decide whether this data is applicable to *in vivo* conditions and what, if any, adjustments should be made. This problem is compounded by the fact that this data is frequently taken from the literature so that it may have been obtained under different experimental conditions. The third problem arises in the interpretation of the results. Where results do not agree with experimental data it may be difficult to determine if the problem lies in the data, the model or in the numerical solution.

One of the first examples of these models was developed by the Cornell group (Schuler 1991, Shu and Schuler 1989). This was a single-cell model of *E.coli* which responded to changes in the carbon source (glucose) concentration and the nitrogen source (ammonium) concentration. This model was able to predict macromolecular

composition, cell size and shape, length of C and D periods, point of initiation and growth rate. The parameters used were based on literature and experimental values. This model which is of a type shown in Figure 1.8, has provided the basis for many other studies. Schuler (1991) has expanded it to form a population model which includes details of plasmid multimerization. Peretti and Bailey (1986) have included in their model of glucose-limited growth of *E.coli* the dependence of transcription on RNA polymerase and the mechanism of translation initiation. This type of model helps us to predict what happens when we introduce a perturbation such as a genetic alteration or extra-chromosomal vector into a cell.

The reaction of a cell to growth on different substrates has been studied by Shu and Shuler (1989) who examined the growth of *E.coli* on a glucose/glutamine/ammonium medium. This study revealed a lack of knowledge about the regulatory control of the TCA cycle.

We can see another example of a structured model in Figure 1.9. This is a model for the growth of *Bacillus subtilis* which also examines cellular differentiation (Jeong and Ataai 1990). Cellular differentiation is investigated by including details of purine nucleotide metabolism which is involved in the initiation of sporulation. This model requires the solution of 35 coupled and non-linear differential equations involving almost 200 parameters and its solution was carried out on a Cray supercomputer.

The difficulties inherent in solving a numerical problem of this size have long been a deterrent in the use of structured models. There are ways of reducing the size of the problem, however, without making too many simplifying assumptions. One of these is by making a quasi-steady state assumption. This involves classifying the reactions in the pathway either as slow reactions or as fast quasi-equilibrium reactions. The system is then modelled as moving very quickly onto a quasi-steady state reaction surface as illustrated in Figure 1.10 (Heinrich *et al* 1976). This method has been used by Schuster *et al* (1988,1989) in their investigation of glycolysis in erythrocytes and enabled a reduction in system size from 20 differential equations to 7. We should note that this quasi-steady state approximation is similar to the reduction of the stoichiometric matrix by fixing variables described in Section 1.1.3.1.2.

While there are drawbacks to this method of modelling as mentioned above, it provides one of the few ways of applying *in vitro* data to *in vivo* systems. For systems to which

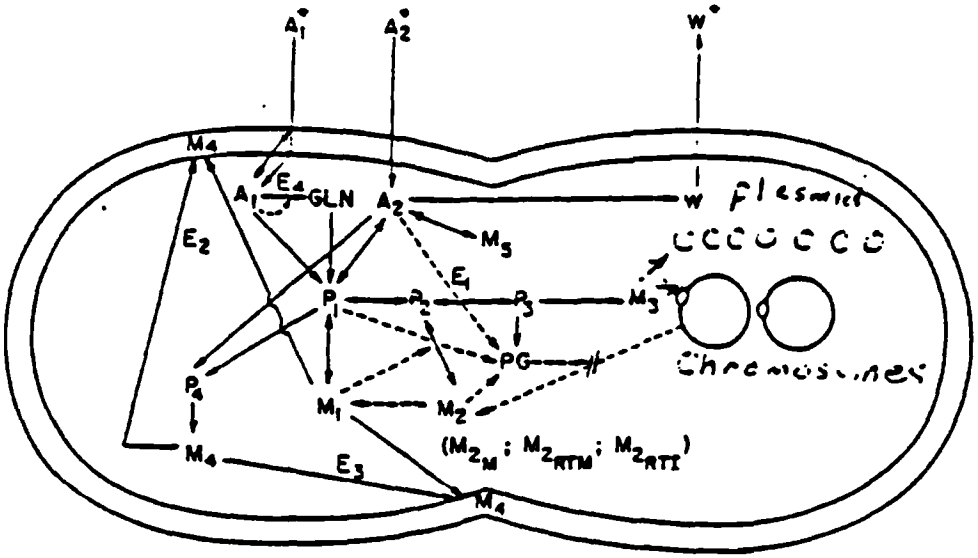


Figure 1.8: Structured model of E.coli

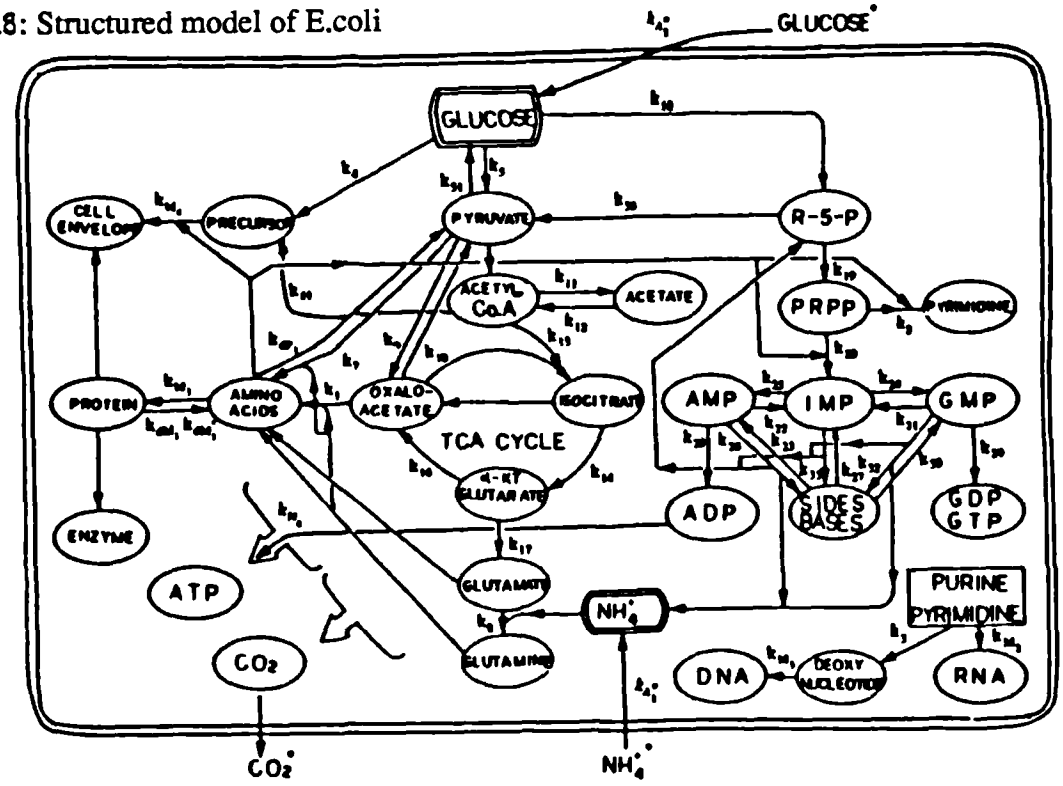


Figure 1.9 : Structured model of Bacillus subtilis

it can be applied it can provide insight into metabolic regulation and interaction, validate or invalidate hypotheses and predict the effect of perturbations on the system.

1.1.3.2 Expert Systems in Metabolic Pathway Engineering:

While great advances have been made in the modelling of metabolic systems by using classical modelling methods, there is a new range of techniques coming of age which could advance our knowledge even further. These methods use a knowledge-based approach and are called expert systems or knowledge based expert systems.

The aim of these approaches is to embed the knowledge of an expert about a system in the form of a series of rules. These rules are of the type that are known and intuitively applied when interpreting data; for example that metabolite concentrations are non-negative, reactions with a negative Gibbs free energy change are thermodynamically favoured and so on. This approach has been used to develop a software system for metabolic pathway synthesis by Seressiotis and Bailey (1988). This system consists of a data base for storing enzyme and substrate descriptions which is used by a search algorithm to identify possible ways to interconvert metabolites which contain carbon.

This approach is an expert system because of the set of rules and restrictions which are programmed in order to guide the search routine. It can only supply the user with a set of possible pathways, however, and experience and knowledge are required to choose between them. In their work, Seressiotis and Bailey were able to identify a route synthesizing L-alanine from pyruvate which does not incorporate the enzyme alanine aminotransferase.

A similar problem was tackled by Mavrovouniotis et al (1990). Their algorithm satisfies a set of stoichiometric constraints by recursively transforming a base-set of pathways. It can thus be used to check the possibility of proposed pathways as well as generating new pathways with desired characteristics. Thus they were able to show that oxaloacetate is a necessary intermediate in all of the pathways from glucose to lysine and that the yield of lysine over glucose cannot exceed 67% in the absence of enzymatic recovery of CO_2.

Because much of our knowledge of biochemical systems is semi-quantitative, the

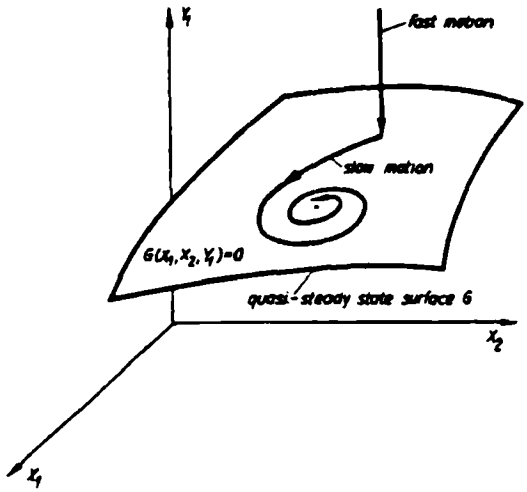


Figure 1.10: Quasi-steady state reaction surface

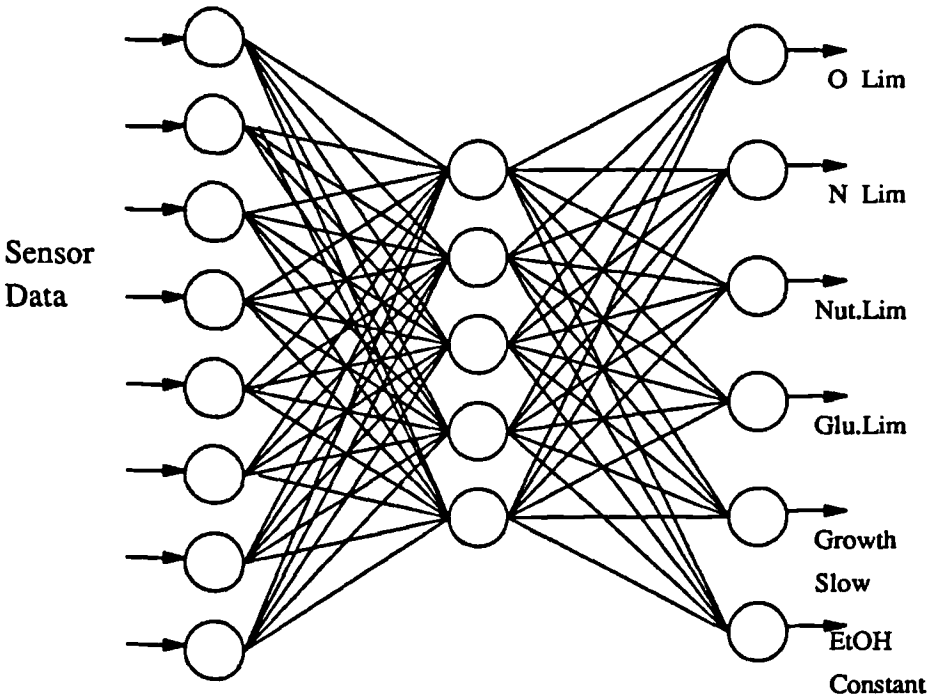


Figure 1.11: A neural network of the type used to examine fermentation data. This shows thow the sensor data is passed through a three layer neural network to provide the output parameters.

formalization of such knowledge is important in the modelling of such systems. This can be approached by order-of-magnitude reasoning (O[M]). Order-of-magnitude reasoning is an artificial intelligence approach based on the formal representation of the relative orders of magnitude of the parameters of a system, through the rigorous definition of relations among quantities (Mavrovouniotis et al 1989). There are seven possible primitive O[M] relationships. O[M] provides an exact interpretation of these relationships and this can be used along with quantitative data to produce inferences. This method has been applied to Michaelis-Menten kinetics, inhibition of enzymatic reactions and identification of rate-limiting steps in biochemical pathways (Mavrovouniotis et al 1989).

Another interesting application of "new" modelling techniques to biochemical systems is that of neural networks. These are dynamic systems composed of highly interconnected layers of simple neuron-like processing elements. A simple neural network is shown in Figure 1.11. The weightings between the neurons are most commonly calculated by back-propagation using a set of training data. This approach could be used to model any part of a cell where we do not wish to go into the detail required by a structured model. This approach has been called a "black-frame" approach. It is illustrated on Figure 1.12 and has been described (without using neural networks) by Schlosser and Bailey 1990.

This is a relatively new area for metabolic modelling and much work remains to be done before it can be easily applied to general systems. The potential of these methods seems great however and with rapid advancements in computer speed and storage progress should be quick.

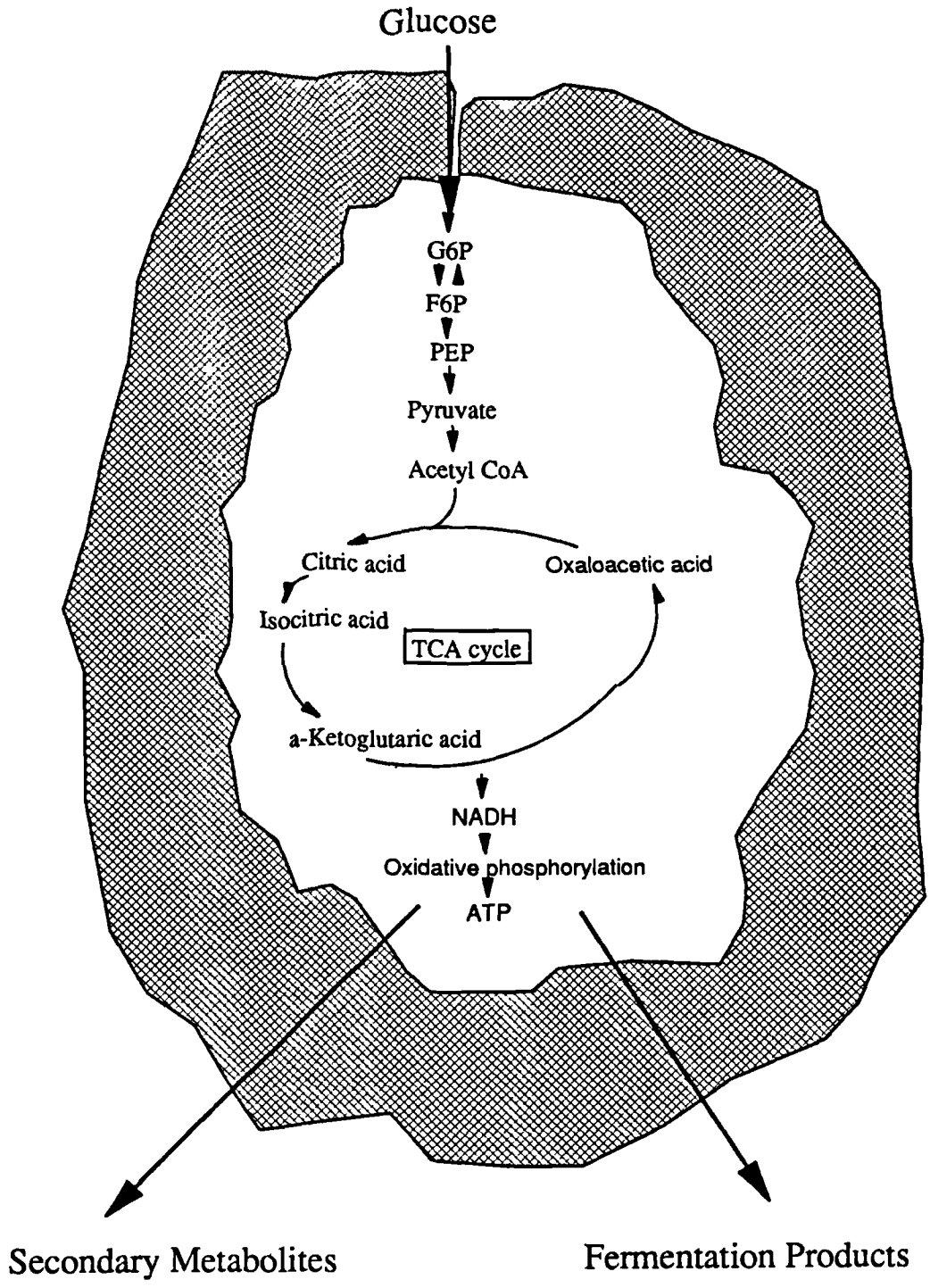


Figure 1.12: Black frame model

1.2 Review of Control Theories

1.2.0 Introduction

The preceeding sections have so far described the different levels of modelling complexity that can be used when addressing a metabolic pathway engineering problem, ranging from phenomenological, through unstructured stoichiometric and structured pathway models, to highly structured metabolic models. The function of these models is to predict or simulate the action of the system.

None of the approaches so far described attempts to quantify the extent to which elements in a system regulate or control the system, or to pick out the most important regulating elements. In an unstructured model all of the elements inside the system are lumped together as a black box whereas in a highly structured model all of the parts of the system are given equal importance. It is often important to be able to identify and quantify the rate-controlling elements in a pathway and it is for this reason that theories of metabolic control have been developed.

The discussion of some of the theories which have been developed to treat the regulation of metabolic systems will begin with metabolic control theory. THis was initially derived for specific cases and then developed for the general case. It presents results which are more immediately applicable in the understanding and treatment of experimental systems This will lead on to the discussion of biochemical systems theory and flux-oriented theory which were initially developed from more fundamental principles.

It will be illustrated that both approaches are complementary and that the appropriate elements of each should be used for the required purposes. The relationship between these theories and other modelling approaches is shown in Figure 1.1.

1.2.1 Metabolic Control Theory:

1.2.1.1 Initial Developments and Their Motivations

Metabolic control theory was established in the mid-seventies with the publication of the papers by Kacser and Burns (1973) and Heinrich and Rapoport (1974). The theory was established in response to the need to quantify the control exerted on a pathway by the enzymes and/or effectors involved in the pathway reactions. It was intended to be a means of combining the information available from the studies of individual enzymes (kinetics, inhibition characteristics, effectors and so on) with that known about the metabolic pathways and their stoichiometry. To avoid the elaborate and non-linear mathematical models usually required to deal with such systems an attempt was made to define the system by parameters which would describe how control of the flux in the pathway was distributed among the enzymes in the pathway.

The approach of Kacser and Burns was to define a range of parameters which they use to describe the effect of perturbations in system variables such as the input, enzyme and intermediate concentrations on the flux and metabolite concentrations. The terms which they use for these parameters are the response coefficient, the controllability coefficient and the sensitivity coefficient. A standard nomenclature was later established (Westerhoff *et al* 1984) in which these are referred to as flux and concentration control coefficients and sensitivity coefficients (see Table 1.2).

Table 1.2: Standard nomenclature of metabolic control theory.

Final Terminology	Earlier Literature
elasticity coefficient of enzyme E_i towards metabolite X_j	elasticity coefficient of E_i effector strength of X_j intrinsic sensitivity sensitization sensitivity amplification amplification factor
flux control coefficient by enzyme E_i on flux J	sensitivity coefficient control strength parameter strength
control coefficient on flux J by pathway substrate S	sensitivity
X_j-concentration control coefficient by enzyme Ei	element of control matrix substrate sensitivity
X_j-concentration control coefficient by pathway substrate S	logarithmic gain

Having defined these parameters and discussed their significance Kacser and Burns then go on to derive a theorem which relates the flux-control coefficients in a straight-line path by considering the changes following a perturbation. By this method they derive the flux control coefficient summation theorem i.e.

$$\sum_{i=1}^{n} C_{E_i}^{J} = 1 \qquad (16)$$

That is, if the regulatory effect of one enzyme on the flux increases then the effect of the other enzymes decreases. This method of proof is useful in that it provides a conceptual means of following what is happening in the system and gives a feeling for the physical interpretation of the result. The disadvantage is that it applies only for very simple linear pathways and additional theorems must be derived when dealing with elements of complex systems such as branched pathways, conserved metabolites or substrate cycles. (see Section 1.2.1.2 below)

Kacser and Burns (1973) also derive the relationship between the flux-control coefficients and the elasticities as expressed by a connectivity theorem

$$\sum_{i=1}^{n} C_{E_i}^{J} . \epsilon_{S_k}^{E_i} = 0 \qquad k = 1,...,m-1 \qquad (17)$$

i.e. the flux control coefficients of highly responsive enzymes are low. We can visualise this in the following way: a reduction in the activity of an enzyme will lead to an increase in the concentration of its substrate and a decrease in the concentration of its product. If the enzyme has a high elasticity towards either the substrate or the product then the change in concentration will lead to an increase in the rate of the reaction catalyzed by the enzyme that compensates almost entirely for the initial reduction in rate. That is, a reduction in the activity of a highly responsive enzyme has only a little effect on the steady-state flux, implying a low flux control coefficient. Summing this effect over the whole pathway we obtain the above result.(Westerhoff et al 1984)

Concentration control coefficient summation and connectivity theorems can be derived in the same way (Heinrich and Rapoport 1974, Westerhoff et al 1984)

Examining the concentration control coefficient summation theorem:

$$\sum_{i=1}^{n} C_{E_i}^{S_j} = 0 \qquad (18)$$

i.e the sum of the concentration control coefficients on a particular metabolite by the different enzymes equals zero, as some enzymes act to decrease the concentration of the metabolite while others increase it.

Similarly the concentration control coefficient connectivity theorem

$$\sum_{i=1}^{n} C_{E_i}^{S_j} \cdot \epsilon_{S_k}^{E_i} = -\delta_{jk} \qquad (19)$$

is analogous to the flux control coefficient connectivity theorem if $s_j \neq s_k$. We can state this theorem as follows: the less responsive enzymes are towards changes in the concentrations of their own substrates and metabolites the more stringent is the control they can exert on a particular metabolite (Westerhoff and Chen 1984).

An example of a simple three enzyme system which illustrates the above theorems is given in Appendix C.2.

1.2.1.2 Extension to Complex Systems:

For simple systems there are some other relationships which can be established, for example, between the elasticity coefficients, flux control coefficients and the flux ratios (Westerhoff et al 1984, Heinrich and Rapoport 1974,1975). However, if we wish to apply metabolic control theory to more complex systems including features such as branched pathways, substrate cycles and conserved metabolites, we need to adapt it slightly. What follows will describe some of the refinements which must be made to the basic theory in order to include features of complex systems. It is by no means a complete list nor does it make the theory applicable to the most general system. It does give an indication of the flexibility of the theory and how it can be adapted given a good understanding of metabolic control theory and metabolic systems. This

requirement is one of the drawbacks of metabolic control theory and frequently makes its implementation difficult for the first-time user.

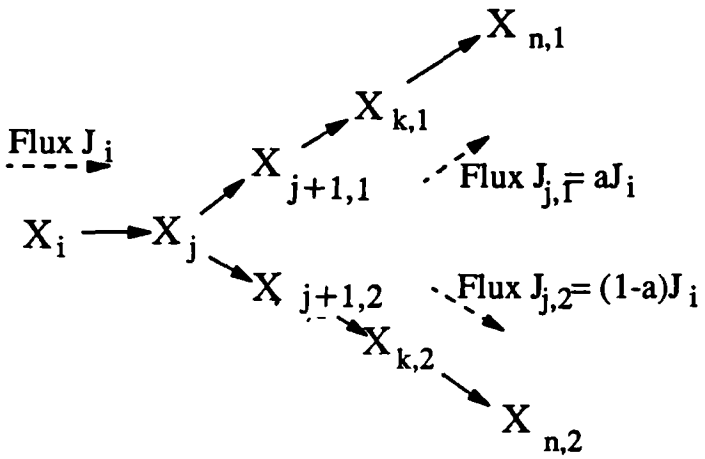


Figure 1.13: System with branched pathways, showing division of fluxes.

For the case of a system with branched pathways as shown in Figure 1.13 an additional theorem was derived by Fell and Sauro (1985, see also Small and Fell 1989) to relate the flux control coefficients at the branch point. This additional equation was shown to be

$$-aJ_i\left(\sum_{h=j,n} C^{i}_{h,2}\right) + J_i(1-a)\left(\sum_{h=j,n} C^{i}_{h,1}\right) = 0 \qquad (20)$$

An equivalent equation for the concentration control coefficients was later derived (Westerhoff and Kell 1987)

$$-aJ_i\left(\sum_{h=j,n} C^{X_j}_{E_{h,2}}\right) + J_i(1-a)\left(\sum_{h=j,n} C^{X_j}_{E_{h,1}}\right) = 0 \qquad (21)$$

Similarly an extra equation for systems with substrate cycles was derived so that solutions for such systems can be obtained (Fell and Sauro 1985)

A slightly different approach must be adopted when dealing with systems containing conserved metabolites. An example of this is in mitochondrial metabolism where the

total of ATP, ADP and AMP is constant. In this case we must use an effective elasticity (Fell and Sauro 1985)

$$\epsilon_{eff,ATP} = (\epsilon_{ATP} - \epsilon_{ADP}\frac{[ATP]}{[ADP]}) \qquad (22)$$

where $\epsilon_{eff,ATP}$ is the elasticity towards ATP under conditions of conservation of ATP and ADP. A combination of the above techniques can be used to deal with the problem of systems involving cyclic pathways with conserved metabolites (Fell and Sauro 1985). The connectivity and summation theorems in Eqn.s 16-19 will not hold unless the assumptions of free enzymes and pooled metabolites are valid.

Metabolic control theory has been used to investigate enzyme-enzyme interactions (Kell and Westerhoff 1990), homologous and heterologous sites (Welch and Keleti 1990), channelled metabolites (Keleti and Ovadi 1988) and high enzyme concentrations (Fell and Sauro 1990). It has also been generalized to conditions of no proportionality between activity and concentration of enzymes (Melendez-Hevia *et al* 1990). Thus we see one of the principal disadvantages of metabolic control theory as it was initially established. Because it is not sufficiently general it must be adapted and tailored to apply it to individual cases. The adaptations required may be only slight but making them requires judgement and a degree of familiarity with the theory. This has a double disadvantage: firstly it will discourage workers in fields ranging from biochemical engineering to molecular biology who are not familiar with metabolic control theory from using it and secondly it makes the writing of software packages utilizing the approach more difficult both from the point of view of program structure and the use of numerical methods.

This problem was alleviated with the extension of the theory to the general case and the description of the system by matrix algebra.

1.2.1.3 Formalization and Development of a Matrix Algebra

In order to solve the difficulties stated above a two-pronged approach was adopted. The first element of this was the generalization (and, necessarily, formalization) of metabolic

control theory (Reder 1988)and the second was thedescriptio nof the system by matrix algebra (Small and Fell 1989, Westerhoff et al 1984, Fell and Sauro 1985).

To generalize metabolic control theory it was necessary to take a "structural" approach. This approach takes the stoichiometric reaction without any expression of the reaction rate as the structure of the model (Reder 1988). Firstly a reaction matrix **N** similar to that in Section 1.1.3.1.2 is set up. This simplifies the problem, using a notation already established in the field.

Having set up this model system Reder uses decomposition of the matrix **N** to establish a set of independent reactions denoted by the matrix $\mathbf{N_R}$ i.e.

$$\mathbf{N} = \mathbf{L}\,\mathbf{N_R} \tag{23}$$

where **N** is the m x r reaction matrix and $\mathbf{N_R}$ is the m_o x r independent reaction matrix (m is the number of species, r is the number of reactions and m_o is the rank of N).

L is an m x m_o matrix of the form

$$\boldsymbol{L} = \begin{bmatrix} 1 & 0 & . & . & . & 0 \\ 0 & 1 & & & & 0 \\ . & . & . & . & . & . \\ 0 & . & . & . & 0 & 1 \\ & & \boldsymbol{L_o} & & & \end{bmatrix} \tag{24}$$

where $\mathbf{L_o}$ is an $(m\text{-}m_o)$ x m_o matrix

This model system and this matrix of independent reactions is then used to express the ideas of metabolic control theory. A steady state flux function **J** and a set of rate vectors **V** are defined and the relationship between a steady state control matrix **Γ**, a steady state flux control matrix **C** and the system variables is established.

The above control matrices differ from those used by other authors in that the coefficients are "simple" and not logarithmic derivatives. The reason given (Reder, 1988) is that the logarithmic steady state flux control matrix has fewer general structural linear relationships between its rows and its columns (for example the summation and connectivity relationships) than has the simple steady state flux control matrix.

The structural properties including summation and connectivity relationships of these simple steady state control matrices are then outlined, using a rigorous mathematical

method. The model differential system described below is similar to that in Equation 3 in Section 1.1.3.1.2 but not at steady state. The form of these relationships is detailed in Appendix C.3.

This treatment is very useful in that it formalizes and makes general a theory which had previously been accused of being set up in an *ad hoc* manner (Savageau *et al* 1986). By stating specifically all assumptions made and using rigid proofs it further validates what was earlier proven in a more intuitive manner. Unfortunately the terminology and methods used are a little inaccessible to those without some background in mathematics. Before it can be used by the average researcher in biotechnology or biochemistry some "translation" is required. This translation could perhaps be in the form of a user-friendly computer program which would convert an input system into the reaction matrix and then deliver the control matrices. We should note that if this method is to be used then some kind of normalization must be chosen in place of the logarithmic normalization used by the other methods.

Even before the formalization of metabolic control theory (Reder 1988) it had been shown that for ease of manipulation and obtaining numerical solutions a matrix representation was required. This was first done to facilitate the expression of flux control coefficients in terms of the elasticities (Fell and Sauro 1985) as will be discussed in the following section. An extension to this theory allowed the calculation of the concentration control coefficients in a similar manner (Sauro *et al* 1987) and determined the response of the flux at the branch points in a pathway. Also in the same paper, Sauro *et al* outlined a set of rules or heuristics for setting up the matrix notation of metabolic control theory applied to a system.

Another work published independently at about the same time proposed a similar set of matrix equations relating the flux and concentration control coefficients in branched pathways to the elasticity coefficients (Westerhoff and Kell 1987)

More useful perhaps is a review of the validity of these matrix methods as applied to complex pathway structures. (Small and Fell 1989). Here the structural approach to metabolic control theory of Reder (1988) is retrospectively applied to the above matrix expressions to test for their validity in complex and general cases. In this way, quite complicated structures such as multiply branched pathways or branched structures with reconvergence can be expressed in matrix notation.

Because of these developments the criticism that metabolic control theory must be developed in an *ad hoc* manner for different systems and revalidated for each one is becoming less justified. The theory has now been proven generally and the time is now fast approaching when it will be amenable to numerical representation and analysis regardless of the complexity of the system under consideration.

Other approaches avoided these initial difficulties by using a power-law approximation for metabolic rates. These were called biochemical systems theory and flux-oriented theory and are discussed in Section 1.2.2.

1.2.1.5 Other Applications

The aim of this section on metabolic control theory is to show how the basic principles can be extended and used to treat many different systems involving transient responses, thermodynamic effects and microbial growth.

The techniques of metabolic pathway engineering have been used in the analysis of transition times by Easterby (1990) and Melendez-Hevia (1990) among others. The aim of these works is to analyze the establishment of a new steady state from a given steady state and the systems response to perturbations while in such a steady state. To this end a temporal control coefficient (Easterby 1990) or transition control coefficient (Melendez-Hevia 1990) was defined:

$$C_{E_i}^{\tau} = \frac{E_i}{\tau}\left(\frac{\partial \tau}{\partial E_i}\right) \tag{25}$$

where τ is the transition time for a steady state established from rest defined by

$$\tau = \sum_{j=1}^{m} [S_j]/J \tag{26}$$

where $[S_j]$ is the concentration of metabolite j and J is the steady state flux to product. Summation and connectivity relationships similar to those in Section 1.2.1.3 have been proven so that the temporal control coefficient can be calculated from the elasticities of

a system in the same way as for flux control coefficients (Torres *et al*, 1994).

The analysis of this transition behaviour is a departure from the other work in metabolic control theory which studies systems at steady states. This kind of work could be very useful in determining which are the parameters that contribute to the instability of a system (Mizraji *et al* 1988). For example it could be used to quantify relationships which have previously been studied only qualitatively, such as that between the pulse of NADH and the onset of oscillations in glycolysis (Riol-Cimas and Melendez-Hevia 1988, Goldbeter and Nicolis 1976).

Some interesting studies have used metabolic control theory to explain the way that variations in enzyme activity segregate in a population (Keightly 1989). Kacser and Burns (1981) have also used it to offer an explanation for the general recessivity of null mutants in diploid organisms which goes as follows: since the flux control coefficient of most enzymes is low, a decrease in their concentration tends to have little effect on the pathway flux so that many rounds of mutations are normally required to produce productive strains.

There has also been a great deal of interest shown in the ways that metabolic control theory can be used to analyze the thermodynamics of metabolic systems. Since living systems are not in equilibrium (the only means by which an organism can attain equilibrium is through death, which is an interesting philosophical point) it is necessary to use non-equilibrium thermodynamics, an area pioneered by Prigogine (1967) and recently reviewed by Rutgers *et al* (1991).

Kell *at al* (1989) have published a review of how microbial growth can be studied by control analysis. In this they discuss the role which metabolic control theory can play in modelling microbial growth and how it compares with other theories such as biochemical systems theory and non-equilibrium thermodynamics.

We can see from the above how broad the range of applications of metabolic control theory can be and with improvements in experimental and computational techniques, this range will expand even further.

1.2.2 Biochemical Systems and Flux-Oriented Theory

1.2.2.1 Derivation of the Power-Law Formalism

One of the earliest attempts to establish an underlying formalism which could be used to describe metabolic systems was made by Savageau in 1969. [Savageau 1969a,b,1970] In these papers a power-law approximation technique is proposed as a description of the dynamic equations governing biochemical systems. This approximation (or formalism, as it is later referred to) was chosen because it seemed to retain the essential non-linear features of such systems, while being amenable to mathematical analysis.

The power-law formalism is analogous to the linear formalism, being based on the first two terms of the Taylor series, but in logarithmic space. We can write the Taylors series for the rate of reaction v_i about an operating point O as

$$\ln v_i(X_1,\ldots,X_n) = \ln v_i(X_{1O},\ldots,X_{nO}) + \sum_{j=1}^{n} \frac{\partial[\ln v_i(X_{1O},\ldots X_{nO})]}{\partial[\ln X_j]} (\ln X_j - \ln X_{jO}) + \ldots \qquad (27)$$

where $X_1,\ldots,X_n$ denote the variables that affect the process such as the concentrations of reactants,enzymes or effectors. Regrouping terms, we can rewrite Eqn. 27 as

$$\ln v_i(X_1,\ldots,X_n) = \ln \alpha_i + g_{i1}\ln X_1 + \cdots + g_{in}\ln X_n \qquad (28)$$

such that the rate can be expressed as a product of power-law functions[Voit and Savageau 1982a, Savageau *et al* 1987a].

$$v_i(X_1,\ldots,X_n) = \alpha_i \prod_{j=1}^{n} X_j^{g_{ij}} \qquad (29)$$

where

$$\alpha_i = \upsilon_{iO} \prod_{j=1}^{n} X_{jO}^{-g_{ij}} \tag{30}$$

$$g_{ij} = \partial(\ln \upsilon_{iO})/\partial(\ln X_j) = (\partial \upsilon_{iO}/\partial X_j)(X_{jO}/\upsilon_{iO}) \tag{31}$$

The expression in Eqn. 29 can be recognised as that normally used to describe chemical kinetics where the g_{ij} exponents are termed the kinetic orders and the α_i coefficients the rate constants. The values of these parameters can be obtained in two ways: either by direct calculation if the rate laws and kinetic constants are known, or from flux data, by plotting in log-log coordinates the reaction rate versus substrate (say) concentration over a range of substrate concentrations close to the operating point. The slope gives the value of g and the value of α can be obtained from the intercept with the y-axis.[Voit and Savageau 1982b]

This mode of parameter calculation has the advantage that the power-law parameters can be calculated from flux data rather than kinetic data although if such data is available it can be used in conjunction with flux data to build a complete picture. It should be remembered that it may not be possible to vary a single concentration inside a cell independently, so that kinetic data may occasionally be essential.

The above formalism underlies most of the control theories proposed for simulating biochemical systems and it is important to remember the principal assumption made in deriving it i.e. that it is only exact for regions close to the operating point and the approximation may not hold for conditions very far from the operating point.[Savageau 1973]

1.2.2.2 Different Levels of Aggregation

1.2.2.2.1 The S-System Variant Used in Biochemical Systems Theory

Having developed this formalism there are several ways in which it can be used to model a biochemical system. The first of these is termed the generalized mass action variant (GMA). This applies the above formalism at the level of individual enzyme

catalyzed reactions [Savageau 1969b]. This has been used by some workers but has not been developed extensively and seems to have limited potential.

At one level up from this, methods differ primarily in the level and type of aggregation of fluxes. The first of these which we will discuss is the S-system variant as applied to the system or part of a system shown in Figure 1.14. This variant has been so named for "its usefulness in simulating the synergistic and saturable properties of biological processes" (Sorribas et al, 1989a). The essential feature of this variant is in the aggregation of those fluxes which tend to increase a given substrate into one net rate law and those which tend to decrease it into another. The power-law formalism explicitly represents each of these net rate laws so that for each substrate (or pool of substrates) X_i we can write the equation

$$dX_i/dt = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}} \qquad i = 1,...,n. \tag{32}$$

There is one equation for each of the n dependent variables and the product is over n+m where m is the number of independent variables.α and β are the rate constants for the net increase and net decrease of X_i respectively and g_{ij} and h_{ij} are the kinetic orders for the net increase and decrease respectively of X_i with respect to variation in X_j.

Even within this variant there is more than one way of aggregating the fluxes in a system. Using as an example the system shown in Figure 1.15 there are two principle strategies which have been termed "reversible" and "irreversible" [Sorribas *et al* 1989b,c].

The reversible strategy involves aggregation of the fluxes with the same sense towards a given metabolite (see Figure 1.15) while the irreversible strategy involves the aggregation of the forward and reverse fluxes through each reaction. Both approaches have been compared and contrasted and while all S-system representations of a given biochemical system exhibit the same steady-state behaviour for the dependent concentrations it has been shown that the reversible strategy is superior with respect to accuracy in predicting steady-state flux and transient responses and in terms of robustness [Sorribas *et al* 1989c].

BIBL. LONDIN. I V.

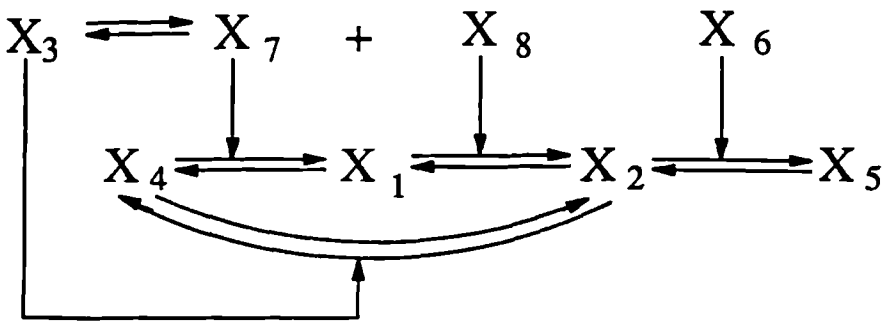
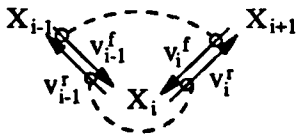


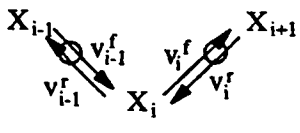
Figure 1.14: Reference system involving enzyme-enzyme interaction and channeling of metabolic flux [Sorribas et al 1989a].

(i) Reversible strategy:



$$dX_i/dt = V_i - V_{-i}$$
$$V_i = v^f_{i-1} + v^r_i$$
$$V_{-i} = v^r_{i+1} + v^f_i$$

(ii) Irreversible strategy:



$$dX_i/dt = V_i - V_{-i}$$
$$V_i = v^f_{i-1} - v^r_{i-1}$$
$$V_{-i} = v^f_i - v^r_i$$

Figure 1.15: Reversible and irreversible strategies

1.2.1.2.2 A Matrix Representation of the S-System Variant.

Having decided on a particular mode of aggregation, in this case the reversible mode we can examine the properties of the steady-state solutions. In steady-state the time derivatives are equal to zero and we can write Eqn. 32 in matrix notation as

$$\mathbf{A}\boldsymbol{y} = \boldsymbol{b} \tag{33}$$

where

$$\begin{aligned} y_i &= \ln X_i \\ b_i &= \ln(\beta_i/\alpha_i) \\ a_{ij} &= g_{ij} - h_{ij} \end{aligned} \tag{34}$$

One of the reasons why this representation is so useful is that it can be immediately determined if the system under consideration has a positive steady-state solution. The condition for this is

$$rank\mathbf{A} = n \tag{35}$$

which, if the system has no external variables can be expressed as

$$|\mathbf{A}| \neq 0 \tag{36}$$

We can seek an explicit solution of the form

$$y_{in} = Ly_{ex} + Mb \tag{37}$$

where y_{in} and y_{ex} are vectors whose elements are logarithms of the internal and external variables and **L** and **M** are matrices obtained from the inversion of Eqn. 33.

From this solution we can obtain the fluxes in the system

$$\ln V = \ln \alpha + Gy \tag{38}$$

where V_i is the net rate of synthesis of X_i, α_i is the corresponding rate constants and the matrix **G** is composed of the g_{ij} kinetic orders.

This solution provides a direct answer to how the dependent variables and fluxes vary with the independent variables but it is useful to calculate the sensitivities of the system

to determine what the rate-controlling parameters are.

The elements of the mxn matrix **L** are defined as logarithmic gain factors [Soribas *et al* 1989a]. These relate the fractional change in any dependent variable to the fractional change in the independent variable responsible for the change i.e.

$$L_{ij} = \frac{\partial \ln X_i}{\partial \ln X_j} = \frac{\partial y_i}{\partial y_j} \tag{39}$$

The elements of the nxn matrix **M** are the sensitivities of the dependent variables with respect to the rate constants. That is, they relate the fractional change in a dependent variable to the fractional change in one of the rate constants.

$$\begin{aligned} S(X_i,\alpha_j) &= \frac{\partial \ln X_i}{\partial \ln \alpha_j} = -\frac{\partial y_i}{\partial b_j} \\ S(X_i,\beta_j) &= \frac{\partial \ln X_i}{\partial \ln \beta_j} = \frac{\partial y_i}{\partial b_j} \end{aligned} \tag{40}$$

$$S(X_i,\beta_j) = -S(X_i,\alpha_j) = M_{ij} \tag{41}$$

We can also define a sensitivity with respect to the kinetic orders i.e.

$$S(y_i,g) = \frac{\partial y_i}{\partial g}\frac{g}{y_i} \tag{42}$$

We can view $S(y_i,g)$ as a weighted average of the sensitivities of the individual L_{ij}'s and M_{ij}'s.

These parameters have also been referred to as control coefficients and an alternative nomenclature has been established (see Table 1.2 [Westerhoff *et al* 1984]) so that confusion can arise. The relationship between sensitivities and control coefficients can be seen more clearly in Appendix C.5 where the derivation of the control theorems described in Section 1.2.1.1 are derived from the power-law formalism.

This approach has been applied to a system involving enzyme-enzyme interactions but

only to the extent that it compares results derived from BST with those derived from Michaelis-Menten rate laws and two other control theories, metabolic control theory and flux-oriented theory which will be discussed below.

1.2.2.3 Flux-Oriented Theory:

The approach taken by Crabtree and Newsholme differs from that outlined above in its emphasis on the fluxes in the system rather than the intermediate metabolites. For that reason it has been named "Flux-Oriented Theory". The reason for concentrating on this aspect of a biochemical system seems to stem from the authors desire to improve our conceptual understanding of how the fluxes and metabolite levels in a system affect each other and change over a given time-period.

This theory was developed independently from the work done by Savageau *et al* but uses the same explicit form of the power-law formalism (called the law of mass action). In the paper in which the bulk of their theory is introduced [Crabtree and Newsholme 1985], the concepts of flux-generating steps i.e. reactions which generate the steady-state flux and flux-transmitting steps, which are the intermediate reactions normally treated as being close to equilibrium. They define a regulatory step as one which communicates with all of the reactions of a pathway and discuss how such regulatory steps are rate-limiting in systems such as the control of postabsorbtive blood glucose levels .

In order to analyze quantitatively such systems they use the response of a given system variable to a given regulator, a quantity which they term the sensitivity s, i.e. the sensitivity of the response Y to the stimulus X is

$$s_Y^X = \frac{(dY/Y)}{(dX/X)} = \frac{d\ln Y}{d\ln X} \qquad (43)$$

Integrating we find

$$Y = k(X)^{s_Y^X} \tag{44}$$

which is a power equation of the kind used in BST by Savageau et al. We should note here that the sensitivities defined in Eqn. 43 used in FOT by Crabtree are not the same as those defined in Eqn.s 39 - 42 used in BST by Savageau. They differ in that they are total differentials as opposed to partial differentials and are in fact similar to the kinetic orders g_{ij} in Eqn. 31. This means that they are global parameters instead of local parameters and that they include the effect of all of the variables in the system rather than just one. We can see this if we write out the full expression for the total differential of Y with respect to X_1 when there are $X_1,\ldots,X_n$ variables in the system.

$$\frac{d\ln Y}{d\ln X_1} = \frac{\partial \ln X_1}{\partial \ln X_1}\cdot\frac{\partial \ln Y}{\partial \ln X_1} + \frac{\partial \ln X_2}{\partial \ln X_1}\cdot\frac{\partial \ln Y}{\partial \ln X_2} + \ldots + \frac{\partial \ln X_n}{\partial \ln X_1}\cdot\frac{\partial \ln Y}{\partial \ln X_n} \tag{45}$$

These s^y_x values are called intrinsic sensitivities and can be combined using a product rule for series reactions or an addition rule for parallel reactions to constitute a net sensitivity to describe all or part of a system.

Using this method, Crabtree *et al* have analyzed the effect of substrate cycling, of systems close to equilibrium and of conserved pools. In this type of analysis the principal aim is to determine how the fluxes affect the sensitivities and vice-versa rather than the steady-state levels of each of the intermediates.

1.2.2.3.1 Treatment of a Substrate Cycle Using Flux-Oriented Theory

An example of this approach can be seen in the treatment of the sensitivity conferred by a substrate cycle. In this system the intrinsic sensitivity for effects of the forward reaction in the cycle is (1+C/J) where C is the rate of cycling and J is the net flux across the cycle [Crabtree and Newsholme 1985].

FOT can be used to determine the conditions under which cycling will increase the sensitivity of the flux J to the regulator X. It can be shown that cycling will increase the sensitivity of the flux J to the regulator X only if the sensitivity of the reaction catalyzed by E_3 to P is greater than the sensitivity of the reverse cycle to P, that is, if P has a greater effect on the forward reaction than the backward reaction to S. A proof for this is given in Appendix C.2.

1.2.2.3.2 Developments and a Matrix Algebra:

This method was later elaborated upon [Crabtree and Newsholme 1988] in order to relate it more closely to the control coefficients used in metabolic control theory. This work displayed a shift in emphasis away from the intrinsic sensitivities and towards the net sensitivities, the s^J_X's. (These are in fact the control coefficients discussed in Section 1.2.1 which have been studied in depth by Westerhoff *et al* among many others.) In this work the slightly unusual notation of X^r is used (to signify an infinitesimal relative change in X) and this makes it rather difficult to make direct comparisons with other approaches i.e.

$$\overset{r}{X} = dX/X = d(LnX) \tag{46}$$

so that for the case of substrate cycling we could write

$$\overset{r}{J_1} - a_1\overset{r}{S_1} - b_1\overset{r}{P} = c_1\overset{r}{X_1} \tag{47}$$

Usually we would have more than one of these equations so that it becomes more convenient to arrange them into a matrix form in order to solve them.

$$\overset{r}{J} = a\overset{r}{S} + b\overset{r}{P} + c\overset{r}{X} \tag{48}$$

$$\begin{pmatrix} 1 & -a & -b \\ . & . & . \\ . & . & . \end{pmatrix} \begin{pmatrix} J \\ S \\ P \\ . \\ . \end{pmatrix} = \begin{pmatrix} cX \\ . \\ . \\ . \end{pmatrix} \tag{49}$$

and so on.

Crabtree and Newsholme (1988) have also showed how this method may be used for more complex branched systems.

1.2.2.4 Comparison of Biochemical Systems Theory and Flux-Oriented Theory:

The treatment of the above system as compared to the BST method gives an indication of how one can concentrate on different elements of a system and provide different insights into it even though the underlying formalism is the same.

Sorribas *et al* have studied this in their comparison of flux-oriented theory and biochemical systems theory although they may have gone too far in their attempts to tailor other theories to the objectives and methods of biochemical systems theory. However they outline some important points at which FOT and BST separate.

The first of these differences is only one of notation but is important both in terms of a conceptual view of the system and its numerical analysis. BST denotes all of its dependent variables by X_i, $i=1,\ldots,n$ and its independent variables by $X_i, i=n+1,\ldots,n+m$ whereas FOT differentiates in its nomenclature between substrates, intermediates, products and enzymes. Both of these notational systems have their advantages and disadvantages. From the point of view of the numerical analysis of a large system the notation used by BST is preferable since in large systems the products of one pathway may be the intermediates, substrates and/or effectors of other pathways and it could be difficult to make the distinctions made by FOT. However the notation in BST is difficult to work with conceptually and it becomes difficult to see how the physical system behaves according to the mathematics.

The second difference is in the aggregation of fluxes. BST strictly aggregates all of those fluxes which go to increase the concentration of a variable into one power-law rate and those which go to decrease the concentration of that variable into another power-law rate. FOT on the other hand stays closer to the traditional biochemical notation of treating each enzyme-catalyzed reaction as an individual rate. The advantages and disadvantages of this approach are similar to those above in that the BST is numerically more convenient but conceptually more difficult. One must be careful not to abandon conceptual mechanisms for the sake of numerical ease. It must be remembered that the aim of modelling a system is not merely to provide a simulation against which to match experimental data but also to provide us with a framework for

the conceptualization of very complex systems.

1.2.3 Implicit Use of the Power-Law Formalism

1.2.3.1 Sensitivity-Based Approaches

Because many workers have been more concerned with predicting the variables whose change will have the greatest effect on a system they have dealt more with the sensitivities of the system rather than the actual values of the variables. As a result their use of a power-law formalism is not obvious although the results achieved are the same as those derived using techniques from biochemical systems theory. This can make it difficult to make direct comparisons between different theories and several authors have attempted to clarify these links.

Each has used the technique of writing down the conservation equations and then differentiating with respect to a given parameter. In the work of Cascante *et al* (1989a,b) the conservation equations are in the form

$$\frac{dS}{dt} = f(S,\alpha,t) \tag{50}$$

$$J_i = \frac{dX_i}{dt} = v_i = g_i(S,\alpha,t) \tag{51}$$

Differentiating with respect to a parameter α_j we obtain

$$\frac{d}{dt}\frac{\delta S}{\delta \alpha_j} = \frac{\delta f}{\delta S}\frac{\delta S}{\delta \alpha_j} + \frac{\delta f}{\delta \alpha_j} \tag{52}$$

and in steady state this implies

$$\frac{\delta S}{\delta \alpha_j} = -\frac{\delta f^{-1}}{\delta S}.\frac{\delta f}{\delta \alpha_j} \tag{53}$$

The authors claim that this contains the same information as Eqn. 33 above which was derived from biochemical systems theory but this is true only if we take δ to indicate the partial differential (a less ambiguous notation would be $\delta_{\alpha j}$)

Sorribas *et al* adopt the same approach applied to a specific system involving enzyme-

enzyme interactions. In their work they write down the steady-state equations before differentiating them (see Eqn.s 43 - 47) in Sorribas *et al*, 1989a and Figure 1.16) so that for the two equations describing the conservation of the interacting enzymes X_9 and X_0 and their product X_3

$$X_9 = X_7 - X_3 \tag{54}$$

$$X_0 = X_8 - X_3 \tag{55}$$

we have, after taking the derivative of the log of the expression

$$\frac{\delta X_9}{\delta X_6}\left(\frac{X_6}{X_9}\right) = \frac{\delta X_9}{\delta X_3}\left(\frac{X_3}{X_9}\right)\frac{\delta X_3}{\delta X_6}\left(\frac{X_6}{X_3}\right) \tag{56}$$

$$\frac{\delta X_0}{\delta X_6}\left(\frac{X_6}{X_0}\right) = \frac{\delta X_0}{\delta X_3}\left(\frac{X_3}{X_0}\right)\frac{\delta X_3}{\delta X_6}\left(\frac{X_6}{X_3}\right) \tag{57}$$

which (using the notation from Eqn. 39) can be written as

$$L_{9,6} = f_{9,3}L_{3,6} \tag{58}$$

$$L_{0,6} = f_{0,3}L_{3,6} \tag{59}$$

Each of the steady-state equations can be differentiated in the same way to give the same results as those obtained from biochemical systems theory (see [Sorribas *et al* 1989b] for details).

1.2.2.2 Development of a Matrix Algebra:

Whereas Sorribas *et al* (1989a,b,c) consider just one specific case, Cascante *et al* (1989a,b) develop a matrix algebra both for a general unbranched system and for

complex systems such as those involving conserved cycles and branched chains. They consider an unbranched pathway which consists of n-1 internal metabolites S_i, h external effectors and n enzymes. In steady state each of the local velocities v_j are equal to the flux J through the system so that for each v_k we can write the control coefficient for one enzyme E_j as

$$C_{E_j}^{J} = \frac{\delta J}{\delta E_j}.\frac{E_j}{v_k}$$

$$= \left(\frac{E_j}{v_k}\right)\left(\frac{\delta v_k}{\delta S_1},\ldots,\frac{\delta v_k}{\delta S_{n-1}}\right)\begin{pmatrix} \frac{\delta S_1}{\delta E_j} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\delta S_{n-1}}{\delta E_j} \end{pmatrix} + \left(\frac{\delta v_k}{\delta E_j}\right)X\left(\frac{E_j}{v_k}\right) \qquad (60)$$

(Note that here δ obviously indicates the partial differential). If we define a sensitivity coefficient as

$$\epsilon_i^j = \frac{\delta v_j / v_j}{\delta S_j / S_i} \qquad (61)$$

we can write Eqn. 60 as

$$C_{E_j}^{J} = \left(\epsilon_{S_1}^{k},\ldots,\epsilon_{S_{n-1}}^{k}\right)\begin{pmatrix} C_{E_j}^{S_1} \\ \cdot \\ \cdot \\ \cdot \\ C_{E_j}^{S_{n-1}} \end{pmatrix} + \left(\epsilon_{E_j}^{k}\right) \qquad (62)$$

In this literature the sensitivity coefficients above have also been called "elasticity coefficients" because they describe the elasticity of the system to changes. If a change in S_i causes little or no change in v_j then the response is inelastic whereas if a large change in v_j occurs then the response is elastic.

Extending this to the set of enzymes j=1,..,n

$$(C^J_{E_j},\ldots C^J_{E_n}) = \left(\epsilon^k_{S_1},\ldots,\epsilon^k_{S_{n-1}}\right)\begin{pmatrix} C^{S_1}_{E_1} & \ldots & C^{S_1}_{E_n} \\ . & & . \\ . & & . \\ . & & . \\ C^{S_{n-1}}_{E_1} & \ldots & C^{S_{n-1}}_{E_n} \end{pmatrix} + \left({}^E\epsilon^k_{E_1},\ldots,{}^E\epsilon^k_{E_n}\right) \tag{63}$$

Rearranging

$$\left(1\ -\epsilon^k_{S_1},\ldots,-\epsilon^k_{S_{n-1}}\right)\begin{pmatrix} C^J_{E_1} & \ldots & C^J_{E_n} \\ C^{S_1}_{E_1} & \ldots & C^{S_1}_{E_n} \\ . & & . \\ . & & . \\ . & & . \\ C^{S_{n-1}}_{E_1} & \ldots & C^{S_{n-1}}_{E_n} \end{pmatrix} = \left({}^E\epsilon^k_{E_1},\ldots,{}^E\epsilon^k_{E_n}\right) \tag{64}$$

and extending this to the n velocities (k=1,...,n) we obtain the equation

$$\boldsymbol{A\ X\ B\ =\ C} \tag{65}$$

where **A**, **B** and **C** are three nxn matrices as follows

$$\boldsymbol{A} = \begin{pmatrix} 1 & -\epsilon^1_{S_1} & \ldots & -\epsilon^1_{S_{n-1}} \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ 1 & -\epsilon^n_{S_1} & \ldots & -\epsilon^n_{S_{n-1}} \end{pmatrix} \tag{66}$$

$$B = \begin{pmatrix} C_{E_1}^{J} & \cdots & C_{E_n}^{J} \\ C_{E_1}^{S_1} & \cdots & C_{E_n}^{S_1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ C_{E_1}^{S_{n-1}} & \cdots & C_{E_n}^{S_{n-1}} \end{pmatrix} \tag{67}$$

$$C = \begin{pmatrix} {}^{E}\epsilon_{E_1}^{1} & \cdots & {}^{E}\epsilon_{E_n}^{1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ {}^{E}\epsilon_{E_1}^{n} & \cdots & {}^{E}\epsilon_{E_n}^{n} \end{pmatrix} \tag{68}$$

If the rate equations are homogenous with respect to the system enzyme concentrations (that is , if the rate laws are proportional to the enzyme concentrations and are independent of each other) then

$${}^{E}\epsilon_{E_i}^{k} = \delta_{ki} \qquad = \begin{cases} 1 & \text{if } k = i \\ 0 & \text{if } k \neq i \end{cases} \tag{69}$$

$$\rightarrow \qquad C = I \tag{70}$$

$$\rightarrow \qquad I = A \times B \tag{71}$$

$$\rightarrow \qquad B = A^{-1} \tag{72}$$

Using this matrix notation the connectivity and summation theorems proven for linear systems in 1.2.1.1 can be shown to hold for general systems (see Appendix C.5 for proofs).

1.2.4 Determination of Control Coefficients and Elasticities

A crucial criterion of the usefulness of metabolic control theory and its applicability to real systems is the ease with which the required parameters i.e. the elasticities and the control coefficients can be obtained. These should be available without having to resort to extensive kinetic or genetic studies of the whole system as this would defeat one of the purposes of metabolic control theory, that is, to direct a cloning procedure so that the flux can be maximized with the minimum number of genetic alterations.

It is the control coefficients which provide the most useful information. They are parameters of the system, rather than of an individual enzyme taking into account the total change in the variable concerned allowing for effects in changes through all the other variables.

This differentiation between the control coefficients as system properties and the elasticities as local properties distinguishes metabolic control theory from biochemical systems theory and also from other sensitivity-based approaches (Kohn *et al* 1979, Kohn and Chiang 1982). It has been emphasized by some authors (Kell and Westerhoff 1986) but has been ignored by others. Indeed the notation used to indicate the control coefficients fluctuates between the total differential d, the infinitesimal change δ and the partial differential ∂. If a coherent literature on metabolic control theory is to be established it will be necessary to standardise these definitions.

Since the control coefficients provide the information on the distribution of control in a system we shall examine various methods of determining them. Indeed we shall see that a principal use of the elasticities is as a means of calculating the control coefficients. The most common ways of determining the control coefficients are by direct calculation, by the use of specific inhibitors, by genetic means or via the elasticities.

1.2.4.1 Direct Calculation:

Many of the early studies identified rate-controlling enzymes by the criterion of the deviation of the mass action ratio from the equilibrium constant (Heinrich *et al* 1977). It was taken as a general rule that enzymes which catalyze reactions far from equilibrium regulate the concentrations of the metabolites. It was shown however that

the deviation of the mass action ratio (Γ) from the equilibrium constant (K_{eq}) is independent of the flux, being a function only of the kinetic parameters of the enzymes concerned and not of the kinetic parameters of any of the preceding enzymes (Heinrich and Rapoport 1974). Because of this, this criterion cannot be used to determine the enzyme(s) which regulate the flux through a pathway, but the ratio Γ/K_{eq} has frequently been used in the expressions for control coefficients, which are derived by direct differentiation.

Table 1.3:

Rate Law	Expression for effector strength
Competitive inhibition: $v = V_{max}\frac{s/K_m}{1+s/K_m+I/K_i}$	$\frac{I/K_i}{1+S/K_m+I/K_i}$
Non-competitive inhibition: $v = V_{max}\frac{S/K_m}{(1+I/K_i)(1+S/K_m)}$	$\frac{I/K_i}{1+I/K_i}$
Allosteric inhibition: $v=V_{max}\frac{S/K_m}{(1+S/K_m)\left(1+\frac{L(1+I/K_i)^n}{(1+S/K_m)^n}\right)}$	$\frac{(nI/K_i)\,L\frac{(1+I/K_i)^n}{(1+S/K_m)^n}}{(1+I/K_i)\left(1+\frac{L(1+I/K_i)^n}{(1+S/K_m)^n}\right)}$

Heinrich *et al* (1977) present expressions for the elasticities (or,in the terminology then used, effector strengths - see Table 1.2) for competitive, non-competitive and allosteric inhibition and give expressions for the control coefficients of a linear enzymatic chain in terms of the relaxation times and the equilibrium constants of the enzymes. (These are presented in Appendix C.6.) For a similar pathway, $S_0 \rightleftharpoons S_1 \rightleftharpoons S_2$, Groen *et al* (1984) give the ratio of the control coefficients in terms of the combined mass action ratio Γ ($= S_2/S_0$) and the combined equilibrium constant K_{eq} (see Appendix C). Derr (1986) presents similar results including the effect of an inhibitor for a similar set of systems.

All of the above results were derived by direct differentiation of the kinetic rate expressions. For more than three steps in a pathway the use of such analytical solutions for the control coefficients becomes difficult. Because of this we must resort to one of the other methods described below.

1.2.4.2 Use of Specific Inhibitors:

This can be regarded as a removal of enzyme from the system. Enzyme-specific inhibitors have frequently been used to determine rate-controlling enzymes on the basis that if, on titrating with the inhibitor, a hyperbolic curve was obtained then the enzyme was rate-controlling whereas if a sigmoidal curve was obtained then it was not. This criterion can be misleading however and is not quantitative (Groen *et al* 1984).

Given an inhibition curve as in Figure 1.16, we can use the expressions in Appendix C.6 to calculate the control coefficients for different types of inhibitors. We can thus see why the shape of the inhibition curve alone is not sufficient to provide even qualitative information on the control exerted by the enzyme since the concentration and nature of the inhibitor must also be taken into account.

This last point was demonstrated by Groen *et al* (1984) in examining the control properties of the adenine nucleotide translocator with respect to oxidative phosphorylation. Because these had been debated in the literature Groen took the approach of estimating the control strength of the adenine nucleotide translocator at different rates of oxidative phosphorylation. By titrating rat-liver mitochondria with the inhibitor carboxytractyloside and isolated rat-liver cells with atractyloside they calculated a minimal control strength of 0.34 for the adenine nucleotide translocator.

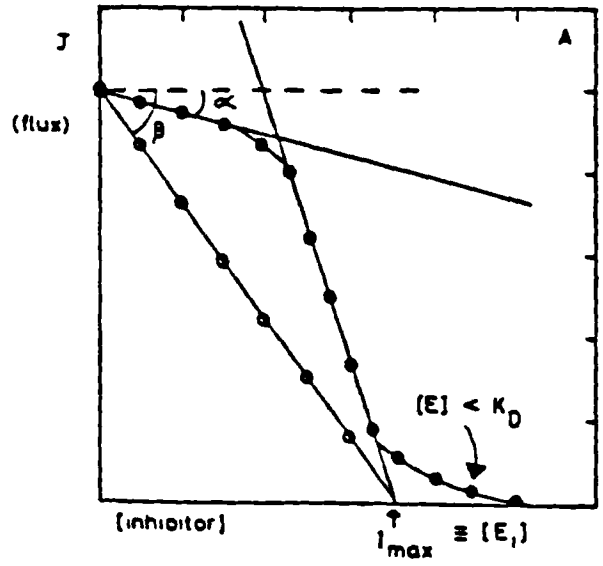


Figure 1.16: Estimation of flux control coefficient of an enzyme by titration with a specific inhibitor. The slope of the line given by α yields the term

$$\left(\frac{1}{J}\cdot\frac{\partial J}{\partial [I]}\right)_{[I]\,=\,0}$$

which is used in the expressions in App.C.6 to calculate the control coefficients.

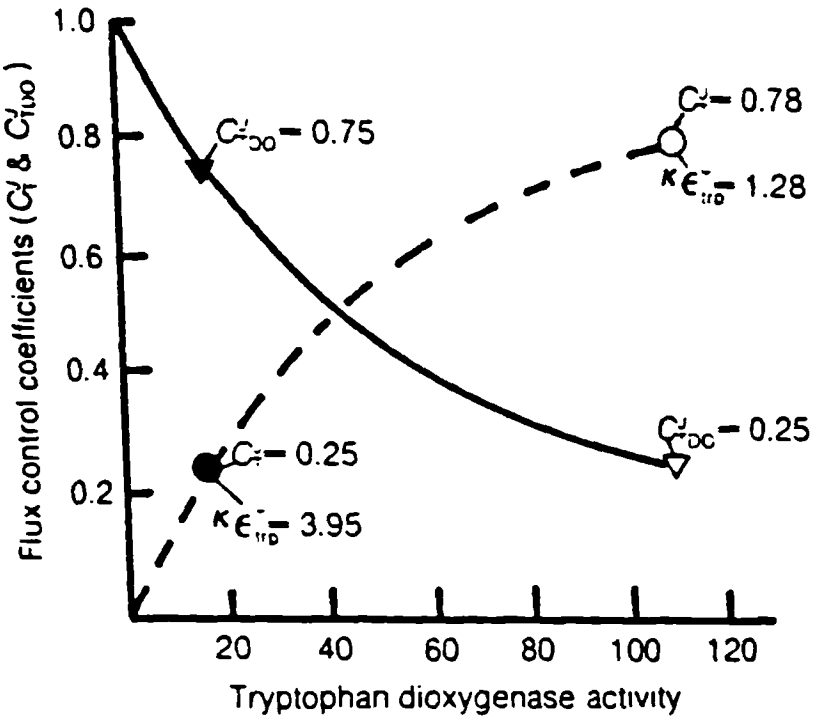


Figure 1.17: The effect of changes in the tryotophan dioxygenase activity on the elasticity coefficient ϵ_{tup} and the flux control coefficients C^{J}T. This shows a non-linear relationship between the control coefficients and tryptophan 2,3-dioxygenase for tryptophan metabolic flux.

In this study the rate of transport of the inhibitor across the membrane was taken into consideration as well as the competitive nature of the inhibitor. A similar study of gluconeogenesis was undertaken, by inhibiting phosphoenolpyruvate carboxylase with 3-mercaptopicolinate. In this case a control strength of only 0.08 was calculated, indicating that the PEP carboxylase exerts minimal control on gluconeogenesis. This result contradicted the conclusions of Rognstad (1979) and Ackerboom (1979) who had used the hyperbolic shape of the inhibition curve as a criterion for control.

The use of inhibitors has its restrictions however. Specific and non-competitive inhibitors are not very abundant although the approach of applying monoclonal antibodies against individual enzymes in permeabilized cells could be used (Kell and Westerhoff 1986). In the use of competitive, non-competitive or mixed-type inhibitors, knowledge of the kinetics and of metabolite concentrations is usually required (as for the adenine nucleotide translocator (Groen *et al* 1984)). In the case of all but irreversible inhibitors we must know the distribution of the inhibitor across any membranes in the system. In the case of irreversible inhibitors, any significant binding of the inhibitor to other proteins will lead to an underestimation of the control strength. Where these restrictions do not apply enzyme-specific inhibitors can provide a convenient method for determination of the control coefficients of a pathway.

1.2.4.3 Addition of Enzyme:

Another method of calculating the control coefficients is by monitoring the change in flux following an increase in enzyme activity or concentration. As with enzyme-specific inhibitors this method cannot be applied to all systems (for example, in the case of membrane-bound enzymes) but when it is applicable it works very effectively.

One of the earliest examples of the calculation of control coefficients by the increase of enzyme activity is the study of bacteriorhodopsin liposome by Westerhoff and Arents (1984). Using an illumination range where the catalytic activity of the bacteriorhodopsin was proportional to the light intensity, they increased the activity of the bacteriorhodopsin by varying the light intensity. They also varied the charge-compensating ion movement by titration with valinomycin and increased proton leakage by adding an uncoupler chlorprimazine. The control coefficients thus found were: 0.9

for the bacteriorhodopsin, 1.0 for the charge-compensating ion movement and -1.1 for the proton leakage reaction. the sum of the control coefficients was therefore 0.8 which is a reasonable verification of the flux control coefficient summation theorem (see Section 1.2.1).

Saltar *et al* (1986) studied the effect of changes in tryptophan dioxygenase activity on the magnitude of the external elasticity coefficient $^{K}\epsilon$ and the flux control coefficients during tryptophan flux in hepatocytes, the results of which are shown in Figure 1.17. We see that as the magnitude of the flux control coefficient for tryptophan dioxygenase decreases with increasing dioxygenase so the magnitude of the tryptophan transport coefficient increases, so that the sum of the two is approximately 1.0 over the range of activity shown. This indicates that the other enzymes of the tryptophan catabolic pathway contribute little to the control of flux over this activity range (Kacser and Porteous 1987).

A similar test of the summation was reported by Rapoport *et al* (1976). Here an equal concentration of purified enzyme was added to ultrasonicated erythrocyte hemolysate in which the ATP level and glycolytic flux had been constant for one hour and the flux control coefficients were as follows: pyruvate kinase 0.0, glyceraldehyde phosphate dehydrogenase -0.08, phosphofructokinase 0.42 and hexokinase 0.33, adding up to 0.67.

It is interesting to compare these results with those from a study by Torres *et al* (1986) of the same system *in vitro*. This study was carried out using rat liver extract which converts glucose into glycerol-3-phosphate. The enzymes fructose biphosphate aldolase, triosephosphate isomerase and glyceraldehyde-3-phosphate dehydrogenase were added to the mixture so that all of the control rested in the first three enzymes of the pathway: hexokinase, glucose-6-phosphate isomerase and phosphofructokinase. The results are given in Table 1.4 with the corresponding figures from Rapoport *et al* (1984) for comparison.

The difference in the above results should alert us to one of the major drawbacks in attempting to establish control coefficients by this method, that is, that we cannot be sure how close our experimental conditions are to that of the *in vivo* system. If the conditions are too different then we may obtain misleading results, as above. It is important to match experimental conditions such as enzyme activity, metabolite

concentration and redox charge as closely as possible to those known to exist *in vivo*.

Table 1.4: Comparison of glycolysis flux control coefficients from two different studies.

Enzyme	CJE (Torres et al)	CJE (Rapoport et al)
Hexokinase	0.77 +/- 0.025	0.69-0.73
glucose-6-phosphate isomerase	0.0 0.0	
Phosphofructokinase	0.24 +/- 0.01	0.31-0.27
Fructose 1,6-biphosphate aldolase	0.0	
Triose phosphate isomerase	0.0	
Glycerol-3-phosphate dehydrogenase	0.0	

1.2.4.4 Genetic Methods:

The problems listed above can be avoided if there exists a means of varying the enzyme concentration from inside the cell. This is the ideal way of establishing the control distribution since the same intracellular and extracellular conditions can be used for analysis as will be used in larger fermentations.

The obvious means of achieving this is by using an adjustable expression vector, for example one containing the *tac* promoter which can be switched on to varying degrees by different levels of IPTG. An example of this has been given by Walsh and Koshland (1985) who put citrate synthase under the control of such a promoter in order to study the extent to which it controls the overall rate of carbon flow through the Krebs cycle. The conclusion that they arrived at was that citrate synthase has little controlling effect during growth on glucose but during growth on acetate almost all the control rests with it i.e. its flux control coefficient is almost 1.0.

An earlier study by Flint *et al* (1980,1981) used this approach by creating heterokaryons of *Neurospora crassa* expressing different amounts of the enzyme of

interest.

There are several pitfalls to be encountered in this method of determining control coefficients. Pleiotropic effects must be avoided and the control should be a null allele of the enzyme of interest. This requires that the genetics of the system must be well-established before this approach can be used, but as more and more knowledge is being obtained about a wide range of systems so should the popularity and applicability of this method increase.

1.2.4.5 Via Elasticities:

As stated above, elasticities are local properties, measuring the degree to which an effector changes the rate of an enzyme while all of the other variables are held constant. Because of this it is possible to measure the elasticities experimentally by incubating the enzyme in conditions as close as possible to those existing in vivo (in terms of concentrations of metabolites, pH and temperature) and then measuring the change in rate upon addition of a small amount of the effector (Kell and Westerhoff 1986). While this does not always ensure that the elasticity obtained will be the same as that *in vivo* (because of the possible absence of allosteric effects *in vitro*, for example) it has been used successfully in the past (Torres *et al* 1986).

If the rate expression for the enzyme is known, then the elasticity can be calculated by differentiation of this expression with respect to the effector considered. This yields expressions such as those in Appendix C.6 (Heinrich *et al* 1977) and these expressions can be used to make rough estimates, even if the enzyme properties are only roughly known. A study of the flux control coefficient varies following different rate equations has been presented by Canela *et al* (1989).

We can use the elasticities thus obtained to yield values for the control coefficients by means of the relationships established in the control theorems (Eqn.s 57 - 60). Groen *et al* (1984) have demonstrated this for a simple two-enzyme pathway of the type

$$S_0 \overset{E_1}{\rightleftharpoons} S_1 \overset{E_2}{\rightleftharpoons} S_2$$

where the relationship has proved to be

$$C_1 = \frac{1}{1 - {}^{v_1}\epsilon_{S_1} / {}^{v_2}\epsilon_{S_1}} \qquad (73)$$

$$C_2 = \frac{1}{1 - {}^{v_2}\epsilon_{S_1} / {}^{v_1}\epsilon_{S_1}} \qquad (74)$$

It is more useful to consider the general case however and since algebraic manipulations of this kind would be tedious we must use matrix methods.

These methods, previously mentioned in Section 1.2.1.3, have been established for both simple (Fell and Sauro 1985) and complex pathways (Sauro *et al* 1987, Westerhoff and Kell 1987, Derr 1986, Fell and Sauro 1985).

The limitations of this method of determining control coefficients, as with the previous forms of matrix representation, are in its lack of generality. To combat this several alternatives to the basic metabolic control theory as established by Kacser and Burns (1973) and Heinrich and Rapoport (1974) have been proposed. We have already mentioned the formalization of the theory by Reder (1986). An equally rigorous method for calculating flux and concentration control coefficients has been derived by Giersch (1988a,b,c) by applying the theorem on implicit functions to the equations defining the steady-state metabolite concentrations (a stoichiometric matrix of the kind described in Section 1.1.3.1). The parallels between this work and that of Reder are obvious.

A different approach based on a non-algebraic diagrammatic method has been proposed by Hofmeyr (1989). In this work "control patterns" showing how a small change in an enzyme activity affects a flux or metabolic pool are drawn directly on the diagram of the metabolic pathway according to certain rules. This method is similar to the King-Altmann method used in enzyme kinetics (Lam 1981). As proposed by Hofmeyr, this method can be used to calculate the control coefficients of systems containing any combination of linear chains, branched pathways, metabolic loops and conserved cycles. It is necessary to work around such features as enzyme-enzyme interactions or channelled pathways, for example by lumping them into one "conversion unit".

A similar topological approach in which the control structure of the pathway is represented by a weighted, directed graph and the control coefficients evaluated in a

heuristic manner from the elasticities is presented by Sen (1990).

These methods have the advantage that they can yield a deeper understanding of the interaction between the local and systemic control properties of metabolic pathways but they are labour intensive and not easily amenable to computer simulation.

It is not always necessary to determine the control coefficients and elasticities for each enzyme in the pathway. Often, topological "functional groups" can be identified and lumped together. This kind of "top-down" approach has been adopted by Brown *et al* (1990). In this work, pathways are conceptually divided in two about an intermediate metabolite and the overall coefficients of the two parts are derived from the overall elasticities of the two parts of the pathway. This approach can be of great help in making the analysis of large systems tractable.

1.3 Conclusions

Reviewing the range of literature on metabolic pathway engineering and the applicable modelling techniques, there are several conclusions which can be drawn.

With all systems some sort of model is proposed. It may be purely phenomenological with no mathematical content as in the case of the engineering of the phenylalanine pathway where the knowledge of the mechanisms of regulation of the system was sufficient to enable engineering of the pathway for maximum phenylalanine production. For example, knowing that it is a tryptophan residue of CMPD which is essential in its feedback repression by phenylalanine enables the removal of this repression by deletion or substitution of the tryptophan residue.

Details of such mechanisms are essential for pathway engineering - we must know where repression occurs before we can remove it - and with a full knowledge of the regulation of a system it may be unnecessary to formulate any other model. Not all systems are so well-documented however and the type of model which can be useful depends on how much data is available on a system as well as on what information we wish to obtain.

The black box system of balancing inputs and outputs over the cell requires only a knowledge of the conservation equations and the elemental composition of the cell. A fermentation equation can then be derived from which can be predicted product yields and also outputs that cannot be directly measured (Papoutsakis 1984, Papoutsakis and Meyer 1985).

If stoichiometric data on the metabolic pathways of the cell is available then we can calculate the fluxes inside the cell and this information can be used in several ways. For example the stoichiometric information on the *Candida lypolytica* system was used to determine which of three metabolic models was the correct one (Aiba and Matsuoka 1979). Flux analysis can also be used to examine carbon flux inside the cell and flux to waste products as well as identifying the junctions at which this flux is controlled (Holms 1986, El-Mansi and Holms 1989).

To quantify the control of flux in a pathway as well as identifying the points at which it occurs, several approaches have been developed. All are based on the concept of sensitivity analysis but they differ in details such as the generality of application,

notational convention, the lumping of the reactions and concentrations in the system and determination of the parameters of the system.

How generally applicable a particular theory is depends on its mathematical rigour. The first metabolic control theories were developed on an almost ad hoc basis and applied to linear systems only whereas more recent theories have been formally developed and apply to totally general systems.

Those theories which are most mathematically rigourous (Reder 1988) tend to be least used in application but this may have more to do with their comparatively recent development rather than any practical difficulties. One advantage of considering simple systems is that the metabolic control theorems which can be derived can be explained in physical terms which give an insight into the distribution of control over flux and concentration in a metabolic system.

Another way in which theories differ is in the type of aggregation used. It has been shown (Voit and Savageau 1987) that aggregation into composite rate laws for net increase or net decrease of each system constituent almost always improves the accuracy of the model. This type of aggregation, however, leads to results on the levels of the metabolite pools rather than individual reaction rates, which may be of more interest to the biochemist.

In determining the parameters of the system, either kinetic or fermentation data can be used in most approaches. If there is full kinetic data on the system (for example as part of a highly structured model) then control theories can be used along with the simulation to predict the distribution of control throughout the system but there may be discrepancies between the results obtained from the *in vitro* data and the *in vivo* system. If fermentation data is used then the control parameters calculated are those pertaining to the *in vivo* system under those conditions. This suggests that while kinetic data for individual enzymes may be used to provide initial predictions, it is necessary to use fermentation data to validate these predictions.

The modelling approaches which will be used in this pathway engineering study are highlighted in Figure 1.1 and are detailed in the following chapter and in Chapter 5.

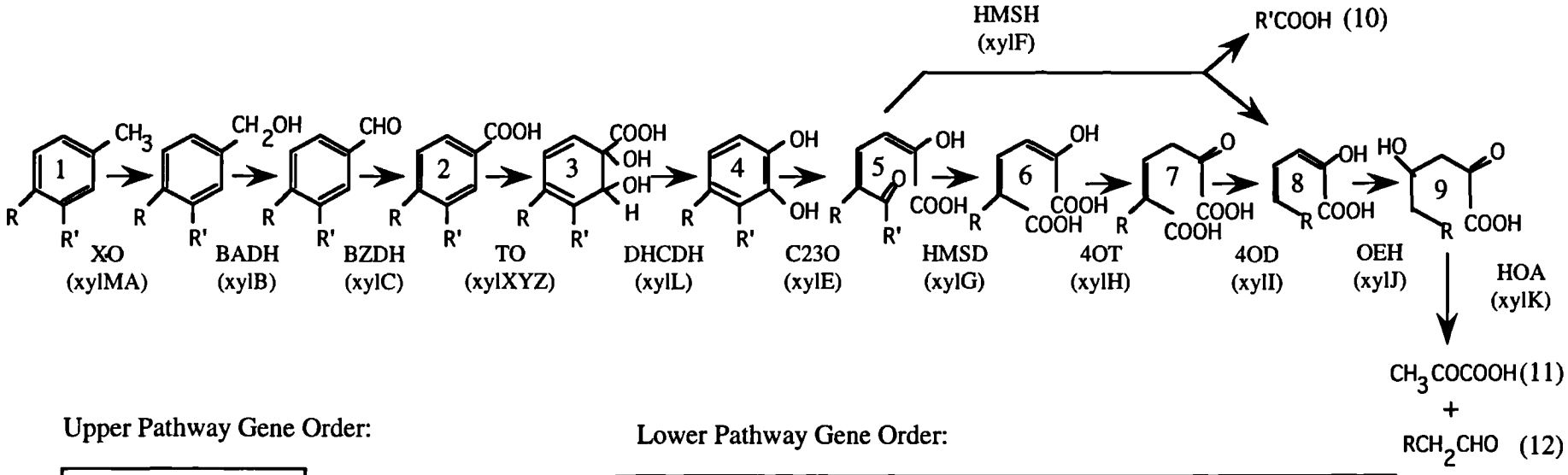
1.3.1 Choice of Test System, the TOL meta-Cleavage Pathway

The TOL meta-cleavage pathway shown in Figure 1.18 was chosen as a test system for several reasons. Firstly, it is of industrial significance owing to its bioremediation capabilities for degrading a range of aromatics and also for its uses in biotransformations for the production of chiral intermediates. Because of these properties, the molecular biology and biochemistry has been extensively studied. What follows is a review of some of the main features of the TOL pathway with respect to its relationship to homologous aromatic degradation pathways, its genetics, the pathway biochemistry, its regulation and how the pathway has been engineered for improved performance.

The TOL system used in this study is the meta-cleavage pathway for the degradation of benzoate to pyruvate and acetaldehyde. It is encoded on the TOL plasmid pWWO which is native to *Pseudomonas putida.* This series of enzymic reactions is one of a family of aromatic degradation systems and degrades substituted benzoates to Krebs cycle intermediates. These aromatic degradation systems are a crucial step in the entry of organic compounds into the carbon cycle. There is an extremely large number of organic molecules which have been created by geological processes and microorganisms have developed a wide range of catabolic pathways in order to use these compounds as growth substrates (Harayama and Timmis, 1989). This is in contrast to the number of organic compounds resulting from biosynthetic pathways, which is relatively limited and so requires only a small number of catabolic pathways for degradation. One essential feature of aromatic degradation pathways is the channelling of structurally different compounds into a small number of catabolic pathways via a central intermediate.

The aromatic ring is very stable and the strategy used by most microorganisms for opening it is by addition of two hydroxyl groups. Once the ring has been doubly hydroxylated it can then be opened by either meta- or ortho-cleavage.

There are five central intermediates: catechol, protocatechuate, homeprotocatachuate, gentisate and homegentisate. The enzyme involved is either a mono- or dioxygenase depending on presence of initial hydroxyl groups. Dioxygenases have a common structure which is that of 4 subunits catalysing reactions. Dehydroxylation occurs



Upper Pathway Gene Order:

(Pu) xylCMABN

Lower Pathway Gene Order:

(Pm) xylXYZLTEGFJQKIH xylS (Ps) (Pr) xylR

R=H, R'=H:

(1) Toluene
(2) Benzoate
(3) 1,2-dihydroxycyclohexa-3,5-diene-carboxylate
(4) Catechol
(5) Hydroxymuconic semialdehyde
(6) Oxalocrotonate (enol)
(7) Oxalocrotonate (keto)
(8) 2-hydroxypent-2,4-dienoate
(9) 4-hydroxy-2-oxovalerate
(10) Formate
(11) Pyruvate
(12) Acetaldehyde

Figure 1.18: The upper and lower TOL pathways

followed by formation of cis-diols. There are two types of fission and two principle catabolic routes. While there is a common strategy among microorganisms for the degradation of aromatic compounds, two means of keeping distinct pathways for different substrates exist: tight substrate specificity and specificity of enzyme induction. Structurally diverse compounds are channelled into a limited number of pathways, yielding a limited number of ring-cleavage substrates which are generally ortho- or para-dihydroxy intermediates. This destabilizes the chemically stable resonant structure of the aromatic ring and enables its subsequent opening. There are two types of fission: ortho- and meta-fission. Both dioxygenases are multimeric proteins, the meta- contains non-haem ferrous irons as prosthetic groups whereas the ortho- contains non-haem ferric irons. After ring-cleavage there are two principle catabolic groups towards the Krebs cycle: ortho-cleavage (or β-ketoadipate) which degrades unsubstituted and haloaromatic compounds to succinate and acetyl CoA and meta-cleavage which degrades alkylaromatics to pyruvate and acetaldehyde.

The pathways are often encoded on conjugative plasmids but some are chromosomal (e.g. the β-ketoadipate). If the TOL plasmid is carried by *P.putida* then the β-ketoadipate pathway is not induced (the β-ketoadipate pathway is induced by cis,cis-muconate, formed by ortho-cleavage of catechol which does not occur).

Catabolic genes are clustered and organised into operons. The order within the operon is identical for isofunctional enzymes and there is strong homology between pathways. However the relative orientation of the two operons and the size of the intervening DNA between the operons vary.

Regulation of TOL and NAH pathways is through modulation of rates of transcription. Substrates of the pathway are inducers of transcription action in conjunction with two regulatory proteins encoded by *xyl*R (upper pathway) and *xyl*S (lower pathway). *Xyl*R requires a sigma factor encoded by *ntr*A and two other factors involved in transcription of nitrogen-regulated promoters can substitute for *xyl*R and stimulate transcription from Pu.

To date, three methods have been used to design microorganisms for increased substrate specificity. These are chemostat selection, *in vivo* genetic transfers (i.e. transfer of DNA via natural genetic processes such as transformation and especially conjugation, for example, natural mutation isolated a TOL *xyl*S*xyl*E mutant able to grow on 4-

ethylbenzoate) and in vitro genetics e.g. creation of a strain capable of degrading mixtures of 3-chlorobenzoate/4-methylbenzoate and 4-chlorobenzoate/4-methlybenzoate by cloning *xyl*IXYZL and an *Alcaligenes* isomerase [Harayama and Timmis, 1989].

1.3.1.1 Historical and evolutionary

In the early 1950s a *Pseudomonas* strain which could degrade aromatic compounds was isolated (Hosowaka, 1952). It was designated *Pseudomonas putida (arvilla) mt-2.* The carbon source used for its selection was m-toluate and benzoate was also reported as a carbon source.

Evidence that the capability for aromatic degradation could be plasmid-encoded was reported independently by three sources in the early 1970s (Nakazawa and Yokota, 1973; Williams and Murray, 1974; Wong and Dunn, 1974) on the basis of a high frequency of conjugal transfer of the ability to utilize specific aromatic compounds and the high frequency of loss of that ability. The plasmid was shown to enable the host to grow on 20 substrates (Table 1.6) and was named pWWO (Worsey et al, 1978). Other plasmids with the same degradation capability have since been discovered, such as pDK1 and pWW53, and all are known as TOL plasmids. However it is the TOL pWWO which is the most studied and which is the source for the meta-pathway used in this study. Initial substrate specificity work an *Pseudomonas arvilla mt-2* was carried out by Murray *et al*, 1972.

1.3.1.2 Relation to Other Aromatic Degradation Pathways

A comparison of gene sequences from TOL PWWO from *P.putida* and those from isofunctional pathways in *Acetinobacter calcoaceticus*, the naphthalene degradative pathway on pVI105 (*P.putida* CF600) and NAH7 from *P.putida* PpG7 suggests that *xyl*XYZL genes diverged from those of *Alcaligenes* 100 to 200 million years ago while the *xyl*TEGFJQKIH diverged from other isofunctional genes 20 to 50 million years. In codons where amino acids were not conserved the substitution rate in the third base was higher than that in synonymous codons and from this it was deduced that both single

and multiple nucleotide substitutions contributed to amino acid-substituting mutations and hence to enzyme evolution [Harayama and Rekik, 1993].

34% of the TOL plasmid shows homology with NAH and SAL DNA. The plasmid pHMT112, carried by *P.putida* ML2, carries genes for the dioxygenase and dehydrogenase involved in the catabolism of benzoate via the ortho-cleavage pathway. These are expressed by the fragment *bed*C1C2BAD in *E.coli* and this fragment has been restriction-mapped [Tan and Fong, 1993].

1.3.1.3 Genetics

The cloning of *Pseudomonas* genes in *E.coli* has been described and reviewed by Nakazawa and Inouye, 1986 and by Mermod *et al*, 1986. Early molecular and functional analysis of the TOL plasmid by transposon mutagenesis determined that the pathway was clustered into two operons, separately regulated by two genes, *xyl*R and *xyl*S [Franklin et al, 1981].

The complete TOL pathway is encoded on a 115 kb transmissible extrachromosomal element, with the genes for the upper and lower parts of the pathway clustered into regulatory blocks and separated by a 10 kb DNA segment. The enzymes encoded by the upper pathway are xylene oxidase (XO), benzyl alcohol dehydrogenase (BADH) and benzaldehyde dehydrogenase (BZDH). The gene order is shown in Figure 1.18.

XO has a relaxed substrate specificity for which a mechanism has been proposed [Harayama et al, 1986a] and enzyme synthesis is positively regulated by *xyl*R. There is 1.7 kb of DNA between the promoter and the *xyl*C gene whose function has not been identified. *xyl*C encodes the 57 kDa BZDH and *xyl*M and *xyl*A encode 35 and 40 kDa proteins which are subunits of XO. BADH is a 40 kDa protein encoded by *xyl*B and the last gene *xyl*D encodes a 52 kDa protein of unknown function. An isofunctional system, *ben*ABCD, has been cloned from *Acetinobacter calcoaceticus* into *E.coli*. *ben*ABC encodes a 53, 19 and 38 kDa protein which comprise the NADH-cytochrome C reductase and terminal oxygenase components of a benzoate-1,2-dioxygenase system. benD encodes the cis-diol dehydrogenase which is a dimer with two identical 31 Kda subunits [Neidle *et al*, 1987].

The meta-cleavage pathway consists of 13 genes and extends over 10 kb. Transposon mutagenesis was used to order and locate eight of the meta-cleavage genes. Some mutations caused what seemed to be non-specific reductions in the activities of several enzymes [Harayama *et al*, 1984].

There are many similarities between the TOL pathway and the isofunctional pathway in *Acetinobacter calcoaceticus*. From the evolutionary relationship between the toluate 1,2-dioxygenase and benzoate 1,2-dioxygenase, the functions of the *xyl*XYZ products were inferred: the *xyl*X and *xly*Y products being two components of an oxygenase and the xylZ product being an NADH-cytochrome c reductase. OEH and 4OD were also found to be associated *in vivo* [Harayama and Rekik, 1990].

Genetic analysis of TO was carried out by complementation analysis using Tn1000 derivatives and nitrosoguanidine-induced derivatives (both defective in TO) and *E.coli rec*A. Four cistrons were identified: the first two comprising *xyl*X, encoding a 57 Kda protein and the third, *xyl*Y encoding a 20 Kda protein [Harayama et al, 1986b].

The *Pseudomonas putida* F1 *tod*C1C2BADE genes were sequenced and cloned into *E.coli* for over-production of their products. Significant homology was observed with isofunctional enzymes from *P.putida* 136R-3 and also with those for biphenyl degradation from *P. pseudoalcaligenes* KF707 [Zylstra and GIbson, 1989].

The gene order for a meta-cleavage pathway from Pseudomonas CF600 capable of utilising phenol, cresol and 3,4-dimethylphenol was determined. There are 15 genes clustered in one operon and these are *dmp*KLMNOPQBCDEFGHI. Enzyme activity assays were used to correlate *dmp*E, G, H and I with known meta-cleavage enzymes and *dmp*F was found to encode acetaldehyde dehydrogenase acylating activity [Shingler et al, 1992]. The activities of the acetaldehyde dehydrogenase (acylating) and the preceding enzyme in the pathway, 4-hydroxy-2-ketovalerate have been shown to be tightly associated. (4-oxalocrotonate decarboxylase and 2-oxopent-4-enoate from the TOL pathway are also tightly bound. *dmp*F and *dmp*G show no sequence homologies.) This metabolic channelling minimizes the toxicity of the aldehyde and maximizes its flux through the reaction, by preventing competing reactions, such as that with ADH [Powlowski et al, 1993].

*xyl*E has been sequenced and its product catechol-2,3-dioxygenase has been purified and

its amino acid sequence determined. It is made up of four identical subunits each containing 1 g atom of iron essential for the activity [Nakai et al, 1983a, Nakai et al, 1983b].

The nucleotide sequence of the promoter region of *xyl*DEGF has been determined, as has the sequence of the *xyl*R and *xyl*S promoter genes [Inouye *et al*, 1984, Inouye *et al*, 1986, Inouye *et al*, 1988].

1.3.1.4 Biochemistry

This section deals with the biochemistry of the principal enzymes involved in the meta-cleavage pathway reactions. These are as follows:

* toluate 1,2-dioxygenase:

TO consists of a flavoprotein, a [2Fe-2S] ferrodoxin (ferrodoxin^{TOL}) and an iron-sulphur protein (ISP^{TOL}). The flavoprotein component of TO (46 kDa) has been purified to homogeneity and shown to accept two electrons from NADH and transfers them to ferrodoxin^{TOL} [Subramanian *et al*, 1981]. The iron-sulphur protein (ISP^{TOL}) is composed of two subunits, with molecular weights of 52.5 kDa and 20.8 kDa [Subramanian *et al*, 1979].The study of the oxidation of indole and indan to 1-indenol and 1-indanol shows that the TO from *P.putida* is capable of benzylic monoxidation as well as dioxygenation. This is also true of naphthalene and may be a characteristic of bacterial multicompenent dioxygenases [Wackett *et al*, 1988].

An investigation of the monohydroxylation of phenol and 2,5-dichlorophenol by TO indicates that only one atom of oxygen is incorporated into the catechol [Spain et al, 1989]. Naphthalene dioxygenase is a multienzyme system consisting of three components: an iron-containing flavoprotein to transfer electrons directly from NADH to cytochrome c, a terminal dioxygenase component which is an iron-sulphur protein (ISP^{NAP}) and a ferodoxin. The terminal dioxygenase has been purified and shown to consist of two subunits with molecular weights of 50 kda and 20 kda, indicating an $\alpha^2\beta^2$ quaternary structure [Ensley and Gibson, 1983]. A 10-fold increase in expression of a 2,3-dihydroxybiphenyl dioxygenase in *E.coli*, *P.putida* and *P.aeruginosa* as compared to the parental strain *P.putida* SU83 has been obtained. This was due to the orientation and transcription read-through from the resistance promoter [Andreyeva *et al*, 1993].

* 1,2-dihydroxycyclohexa-3,5-diene-1-carboxylate dehydrogenase:

DHCDH was shown to be made up of four identical subunits, each of weight 24 kDa. The proposed mechanism of the reaction is dehydrogenation of DHB to an unstable β-ketoacid which decarboxylates to form catechol. K_m values for the reaction were determined to be 0.2 for DHB and 0.15 for NAD^+ [Reiner, 1972].

* 2-hydroxymuconic semialdehyde hydrolase (HMSH):

The meta-cleavage pathway displays a relaxed substrate specificity which is not found in the β-ketoadipate pathway, which cannot metabolise compounds which give rise to alkylcatechol intermediates. While most of the meta-cleavage pathway enzymes have a broad substrate specificity, the enzymes on the divergent branches of the pathway are exceptions and are substrate specific. In theory there are two possible mechanisms for the degradation of 2-hydroxymuconic semialdehyde, one involving 2-hydroxymuconic semialdehyde hydrolase and the other involving 2-hydroxymuconic semialdehyde dehydrogenase. HMSH is most active against 3 substituted catechols and catechol and 4-methylcatechol are almost exclusively reacted upon by the dehydrogenase.

HMSH is encoded by the *xylF* gene which has been mapped to within a region of about 1kb (Inouye et al 1981). It has been purified to homogeneity and shown to have an M_r of 65000. It comprises two subunits of equal size which are assumed to be identical (Duggleby and Williams, 1986). The amino acid composition indicates each subunit as consisting of nearly 300 residues, indicating a structural gene of about 0.9 kb which is in agreement with genetic work.

Only linear chain alkylcatechols and 4-chlorocatechol can be degraded by this enzyme and the reaction proceeds with varying levels of activity which occurs as a result of different V_{max} rather than K_m (see Table 1.5).

* 2-hydroxymuconic semialdehyde dehydrogenase

2-hydroxymuconic semialdehyde dehydrogenase is the other side of the branch in the pathway. With benzoate and p-toluate as substrates for the pathway, its activity is much greater than that of the dehydratase branch and as a result benzoate and p toluate are degraded preferentially by this branch (Harayama et al, 1987). m-toluate is degraded exclusively by the hydratase branch. This branch is energetically favourable as one NADH is produced for each benzoate degraded. The role of the divergent branches was investigated in 1987 and it showed m-toluate to be dissimilated exclusively by the

hydrolytic branch and p-toluate and benzoate by the dehydrogenase branch. Pathway selection is made by HMSD.

Table 1.4: Kinetic parameters for HMSH with various substrates:

Substrate	K_m (mmol)	V_{max} (mmol/hr)
catechol	30	6.9
3-methylcatechol	36	100
4-methylcatechol	10	3.2
3-ethylcatechol	15	34
4-ethylcatechol	22	6.7
3-propylcatechol	5.0	2.4
4-propylcatechol	23	0.5
3-allylcatechol	37	30
4-chlorocatechol	122	2.3

* 4-oxalocrotonate tautomerase

Isomerization of oxalocrotonate occurs spontaneously in vitro but must be enzyme-catalysed in vivo by 4-oxalocrotonate tautomerase. This enzyme catalyzes the enolization of oxalocrotonate from its keto form. This is a reaction which occurs spontaneously in vitro, but whose rate is very low in vivo (Harayama et al, 1987). It is a tetramer of four subunits of 7.5 Kda.

* 4-oxalocrotonate decarboxylase

Relitively little work has been done on this enzyme to date.

* 2-hydroxpent-2,4-dienoate hydratase

This enzyme has been purified almost to homogeneity and has been shown to have a molecular weight of 287,000 . Treatment with SDS suggests that it comprises 10 subunits of molecular weight 28,000. 2-hydroxpent-2,4-dienoate hydratase exhibits a narrow substrate specificity and is stereospecific, hydrogenating only the cis isomer. A mechanism has been proposed for this reaction, based upon the requirement for an enolizable carbonyl group and the stimulation in activity caused by Mg^{2+} (Collinsworth

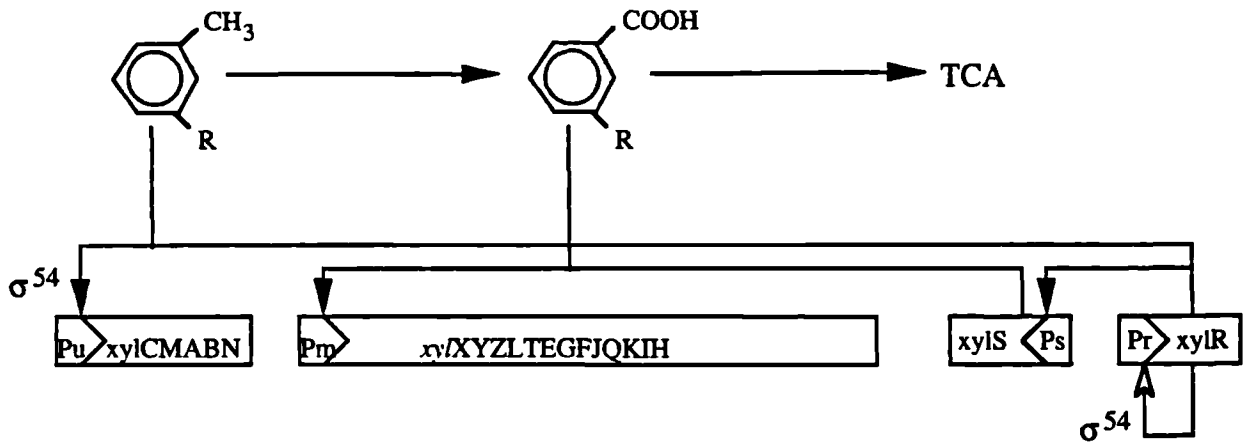
et al, 1973).

* 4-hydroxy-2-oxovalerate aldolase

A similar mechanism for the aldolase reaction based upon the requirement for an enolizable carbonyl group and the stimulation in activity caused by Mg^{2+} (Collinsworth et al, 1973) has been proposed.

1.3.1.5 Regulation

Regulation of TOL and NAH pathways is through modulation of rates of initiation of transcription. Substrates of the pathway are inducers of transcription action in conjunction with two regulatory proteins encoded by *xylR* (upper pathway) and *xylS* (lower pathway). *xylR* requires a sigma factor encoded by *ntrA* and two other factors involved in transcription of nitrogen regulated promoters can substitute for *xylR* and stimulate transcription from Pu [Harayama and Timmis, 1986].



ıre 1.19: This figure shows the gene order of the upper and lower pathways of the TOL plasmid and egulatory mechanisms of the xlyR and xylS genes.. The solid arrows indicate positive regulation ı the required effectors while the open arrow denotes repression.

Figure 1.19 shows the current model for the regulatory cascade which controls the transcription of the TOL genes [Harayama and Timmis, 1989]. Transcription of the upper operon originates from the Pu promoter and is initiated by the *xylR* protein when it is activated by pathway substrates such as m-xylene and toluene and other structural analogues in combination with the σ^{54}-containing form of RNA polymerase. The *xylR*

product seems to regulate its own transcription (indicated by the open arrow in Figure 1.19). Transcription of the lower pathway is initiated when an excess of *xyl*S or *xyl*S bound to its effectors which are substrates of the meta-cleavage pathway activates the promoter Pm.

The promoters Pu and Ps are σ^{54} and IHF-dependent. A directly repeated sequence 5'-TGCAA PuAAPu-PyGGNTA-3', separated by a 6 bp inter-repeat region and overlapping the -35 region of the Pm promoter has been identified as the binding site of the xylS protein [Kessler *et al*, 1993].

The site of *xyl*R interaction has been identified as an inverted sequence repeat homologous to that recognised by *xyl*R in the upper pathway promoter Pu and lying in the region between -133 bp and -207 bp. The *xyl*R protein appears to recognise and bind cooperatively to two halves of the motif in the Ps promoter upstream region. While IHF was found to bind to Ps, it is not essential for its activation (this has been found in other T-rich, σ^{54}-dependent promoters [Holtel *et al*, 1992]. Additional work on the *xyl*R promoter region, studying the positional requirements for the three protein binding sites of Pu [URS for XylR, IHF site for IHF and the -12 to -24 region for the σ^{54}-E complex] by means of point mutations. Two full helix turns did not reduce activity whereas three resulted in a 90% drop [Abril and Ramos, 1993].

A regulatory sequence has been located upstream of the *xyl*S promoter (between 145 and 188 bp upstream of the transcription start site) by deletion analysis. This is a common feature with other σ^{54} dependent promoters characterized to date which all require activator proteins binding to sequences approximately 100 bp upstream of the promoter. This UPS consists of two functional regions and contains a palindromic sequence, suggesting binding in an oligomeric form of the XylR protein. The DNA loop structure proposed for activation of the upper pathway operon [Inouye *et al*, 1990] seems to be applicable to the activation of xylS by the XylR protein in the presence of an effector. It has been postulated that the IHF induces DNA binding and thus facilitates and stabilizes the loop structure in the region between the UPS and the promoter of *xyl*R. There is no IHF binding site between the UPS and the promoter of *xyl*S and IHF mutants show no difference in *xyl*S activity so it appears that a factor other than IHF stabilizes the loop structure [Gomada *et al*, 1992].

Several xylS mutants have been isolated which recognise as effectors benzoate

analogues that are not effectors for the wild-type XylS protein and the characteristics of these have suggested that the effector binding pocket of the XylS protein may be composed of two or more non-contiguous segments of its primary structure [Ramos *et al*, 1990]. Activation by its effectors of the XylS protein is dependent on intramolecular interactions between the C- and N-terminal regions. These domain interactions have been deduced from intraallelic dominance in double mutants on *P.putida* [Michan *et al*, 1992a]. Arg-41 has been identified as a critical residue in *xyl*S for effector stimulation of transcriptional activity [Michan et al, 1992b].

A locus on the *P.putida* chromosome can substitute for *xyl*S in initiation of gene expression from the OP2 lower-pathway operator-promoter region in the presence of benzoate which acts as an inducer. A putative benzoate regulatory gene, *ben*R, has been identified. xylS also substitutes for *ben*R in regulating expression of cloned chromosomal genes. While homology exists between *ben*ABC and *xyl*XYZ, none exists between *ben*R and *xyl*S [Jeffrey *et al*, 1992].

XylS protein mutants have been isolated which constitutively switch on the Pm promoter. These have mutations at their C-terminal ends which are adjacent to α-helix-α-helix domains likely to be involved in DNA binding [Zhou *et al*, 1990].

1.3.1.2 Engineering Microorganisms for Improved Performance

One way of prolonging the life of cultures capable of biocontrol and plant growth-promotion as well as degrading xenobiotics is to enable them to utilize a nutrient which is then supplied to the environment. This has been done by enabling *P.putida* R20 (an antagonist of Pythium ultimum and so reduces rot of sugar beet seed) to grow on salicylate, by the addition of NAH7 [Colbert *et al*, 1993].

A bioreactor for the degradation of benzene, toluene and xylene was developed using silicone tubing for controlled release of substrate and a mathematical analysis carried out [Choi et al, 1992, Lee *et al*. 1993].

Trichloroethylene was degraded by *E.coli* JM109 containing the genes *tod*C1C2BA. A linear rate of degradation was observed whereas in a *P.putida* F1 mutant, the initial rate was much higher and decreased over time [Zylstra *et al*, 1989].

The cis-diols produced as intermediates by the TOL degradation pathway can be used

as building blocks for polyphenylene synthesis or as reactants for the organic synthesis of fine chemicals. The production of cis-diols by TO in *P.putida* and *E.coli* was studied using ^{14}C-labelled benzoate. and it was shown that the presence of a methyl group in the ortho-position relative to the carboxyl strongly interferes with TO activity. Also, the accumulation of cis-diol was much higher in *P.putida* that in *E.coli* [Wubbolts and Timmis, 1990].

Trichloroethylene can be degraded by several aerobic bacteria, some of which have toluene degradation abilities. The products of oxidation of TCE by TO have been shown to be toxic to *P.putida* F1 by the alkylation of cellular molecules, though the microorganism can recover by the turnover and resynthesis of these molecules [Wackett and Householder, 1989].

Several bacterial dioxygenases are capable of the oxygenation of indole to cis-indole -2,3-dihydrodiol. In the case of XO this occurs by direct hydroxylation at C-3. This product spontaneously dehydrates to form indoxyl which in turn oxidises to form indigo. (Indole is formed from tryptophan by tryptophanase) [Mermod *et al*, 1986]. Genes encoding toluene dioxygenase, toluene cis-glycol and catechol-2,3-dioxygenase from *P.putida* NCIB 11767 were cloned into *E.coli* HB101 and were observed to produce indigo and a variety of other coloured products [Stephens *et al*, 1989].

When organisms which can degrade a range of aromatic compound are exposed to mixtures of chloro- and methyl-aromatics, both orth- and meta- pathways are induced to degrade the mixture (chloro- via ortho- and methyl via meta). This leads to non-productive misrouting of the substituted catechols produced. This can be overcome either by restructuring existing pathways or by the constructing of new pathways via the combination of different enzymes from different organisms to give the desired properties. By combining five discrete pathway segments with the desired enzymatic activities a strain was developed which was capable of degrading mixtures of 3CB, 4CB, 4MB, 4CP and 4MP without non-productive misrouting of intermediates [Rojo *et al*, 1987].

Pseudomonas sp. T-12 can degrade many 3-fluoro-substituted benzoates with either defluorinated catechol and inorganic fluoride products or catechol products with the fluorine substituent. The steric size of the C-1 substituent was found to affect the ratio of defluorinated to fluorinated catechols formed and a mechanism for the reaction

involving TO has been proposed [Renganathan, 1989].

Reasons why substituted aromatics or xenobiotics may not be degraded by an organism are: toxicity, failure to be taken up by the bacteria, failure to act as an effector of regulators of transcription of catabolic operons, and failure to be degraded by the pathway enzymes.

The reason why 4EB is not degraded by the TOL plasmid is due to its failure to induce transcription of the pathway enzymes and the fact that one of the pathway intermediates inactivates C23O. (These two points were identified as crucial to expanding the substrate specificity of the TOL pathway) Mutants were selected to overcome these deficiencies and then combined to redesign the pathway [Ramos *et al*, 1987].

The meta-cleavage pathway has been cloned into *Rhodobacter sphaeroides*, and the enzymes were expressed at a high level. However the cells were unable to grow with m-toluate as sole carbon source, probably due to an inhibitory effect of one of the intermediates [Calero et al, 1989].

Cloned genes can be used to extend the range of substrates which can be degraded by an organism. By cloning the genes, *xyl*D, *xyl*L and *nah*G into *Pseudomonas* sp. B13 its catabolic range was expanded to include 4-CB, 3-CB, and 3,5-CB, salicylate, 3-, 4- and 5-salicylate [Lehrbach *et al*, 1984]. The activity of the upper pathway enzymes XO, BADH and BZDH were tested with chlorotoluenes as substrates. Using this information, a hybrid strain was constructed by introducing TOL into the 3-CB-degrading strain B13 or JH320. Steric effects were observed for all three enzymes with substituents in the ortho position bigger than fluorine [Brinkmann and Reinekke, 1992]. Also, the upper pathway enzymes will accept nitroaromatics as substrates though *xyl*R does not recognise them as effectors [Delgado *et al*, 1992].

Several issues involved in using genetically engineered microorganisms for combatting pollution in the environment were identified: survival of the organism, stable maintenance of genome, transfer of genetic material to other microorganisms, spread beyond the target environment and effect on the ecosystem. Two *Pseudomonas* strains, one with the ortho-cleavage pathway for 3CB and 4MB and one with a TOL pathway for 4EB were introduced into an activated sludge microcosm. These protected the indigenous sludge microorganisms from shock loading of these substrates. Horizontal in situ of plasmid DNA was readily observed [Nuβlien *et al*, 1992].

1.3.1.7 Stability of TOL plasmid.

On batch and continuous cultures of *P.putida* on toluene, complete inhibition of growth is observed at greater than 1.4 g/l of toluene. Chemostat cultures can be maintained at steady state for several months [Vecht *et al*, 1988]. The kinetics of plasmid loss of *P.putida* MT15 in potassium-limited chemostat culture were shown to be dependent on the growth rate advantage endowed by the plasmid i.e under glucose limitation there was a stable mixed culture while with excess glucose, 100% plasmid loss was exhibited [Stephens and Dalton, 1988].

TOL plasmid in chemostat culture under succinate-, sulphate-, ammonia- and phosphate-limitation and at different dilution rates has been studied. There was a persistence of TOL+ plasmid-containing cells under phosphate-limiting conditions and it was found that dilution rate had a strong influence under carbon- and energy limitation. Mutants were found containing partially deleted plasmid leading to theories about other functions of the TOL plasmid [Duetz and Van Andel, 1991].

During growth of TOL+ *P.putida* in pH controlled culture, TOL- strains had a 15% growth advantage which led to their taking over the culture. The reason for this was proposed to be an inhibitory intermediate of the meta-cleavage pathway such as hydroxymuconic semialdehyde.

Table 1.6 : List of TOL substrates:

toluene	benzaldehyde
m-xylene(1,3-dimethylbenzene)	m-tolualdehyde (3-methylbenzaldehyde)
p-xylene(1,4-dimethylbenzene)	p-tolualdehyde (4-methylbenzaldehyde)
pseudocumene(1,2,4-trimethylbenzene)	3,4-dimethylbenzaldehyde
3-ethyltoluene	3-ethylbenzaldehyde
benzyl alcohol	benzoate
m-methylbenzylalcohol(3-methylbenzyl alcohol)	m-toluate (3-methylbenzoate)
p-methylbenzyl alcohol (4-methylbenzyl alcohol)	p-toluate (4-methylbenzoate)
3,4-dimethylbenzyl alcohol	3,4-dimethylbenzoate
3-ethylbenzyl alcohol	3-ethylbenzoate

2. SIMULATION AND OPTIMIZATION OF METABOLIC PATHWAYS

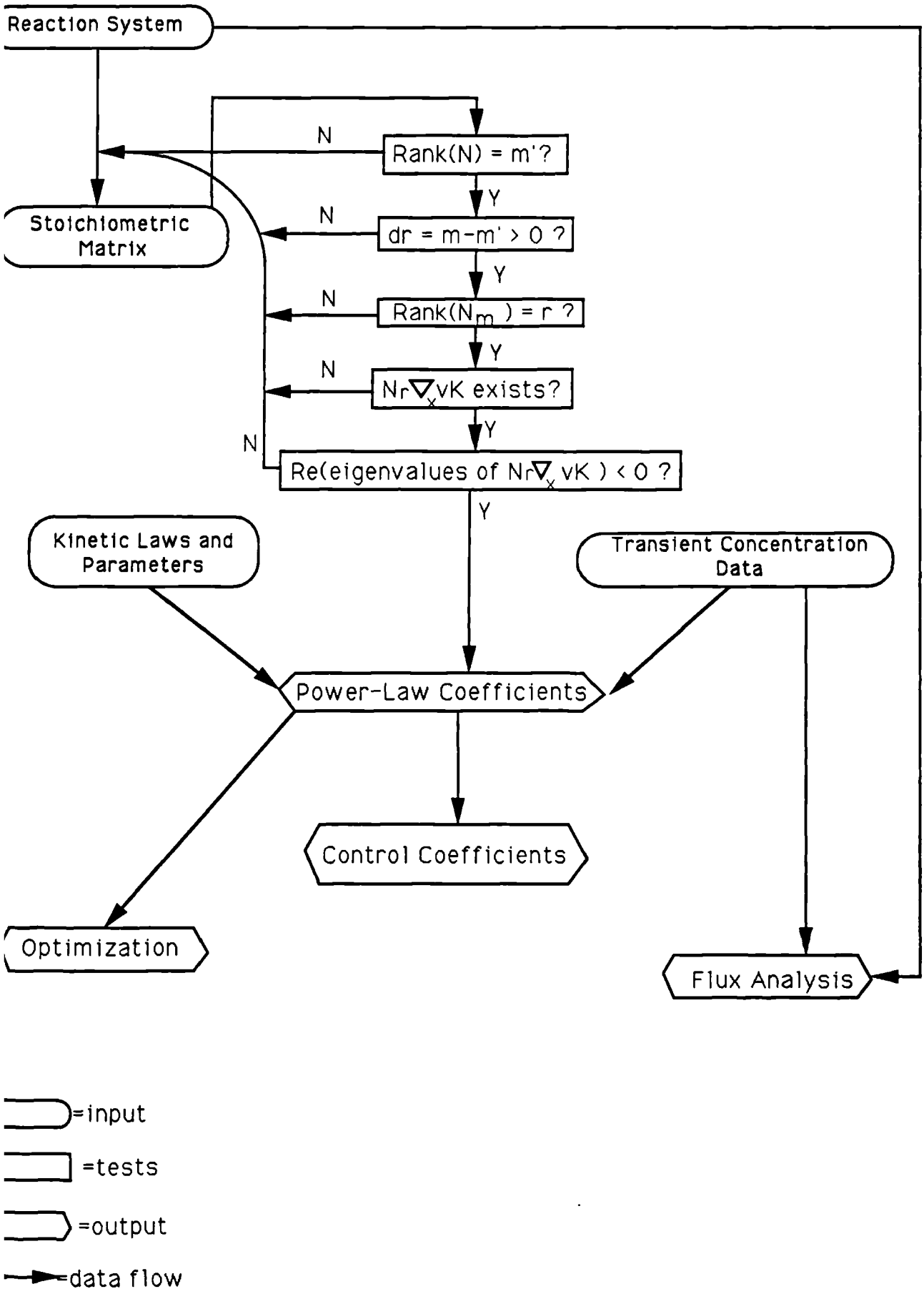
2.0 Introduction

In this chapter, a procedure is outlined for simulating and optimizing a metabolic reaction network. A range of software has been developed for the implementation of this approach with respect to general metabolic systems and this is briefly described. Section 2.1.1 describes how the reaction system matrix is set up and the checks which must be made to validate it and the system steady state. Section 2.2 models the action of the system about this steady state using a technique called biochemical systems theory. This theory uses a power-law function to model the system reactions rather than the more usual linear or Michaelis-Menten expressions. This formalism was chosen because it seemed to retain the essential non-linear features of metabolic systems while being amenable to mathematical analysis [Savageau, 1969a,b, 1970]. Section 2.3 presents the formulation and solution of the optimization problem presented by the system. The method for calculating the control coefficients of a metabolic system from transient concentration data is described in Section 2.4 and methods for sensitivity analysis developed in Section 2.5. These methods have been used in a modelling strategy which is outlined in Figure 2.1. Section 2.6 presents the results of the above methods and software when applied to the TOL meta-cleavage pathway.

2.1. Metabolic Modelling

2.1.0 Introduction

An essential feature of a model is that it establishes a mathematical framework for experimental data. This can be of several forms depending on the function of the model: it can be one that simply arranges experimental data in a logical way so that the



e 2.1: Flow-chart for the modelling system used.

important correlations are emphasised, or it can be one which enables the use of data from individual *in vitro* studies in examination of the overall system behaviour. In any case the model must address two issues: the behaviour of the individual enzymic reactions, and the interrelation of all of the reactions in the system. The individual enzymic reactions can be modelled using different levels of complexity. They can be described mechanistically using expressions of the King-Altman variety, or more simply with a pseudo-steady state model using an expression such as the Michaelis-Menten equation.

Before any system can be modelled it is necessary to define the system boundaries and state which are the exchangeable and non-exchangeable parameters. This helps outline what the modelling assumptions are to be. It is essential to be clear if and when assumptions about a metabolic system are made so if there are discrepancies between the model predictions and experiment, the assumptions can be reviewed. In this case, the system under investigation is the TOL meta-cleavage pathway, whose biochemistry and genetics and regulation are described in Section 1.3.1. The 13 genes of the lower pathway under a *tac* promoter are transformed into *E.coli* JM107 as described later in Section 3.1. The exchangeable parameters are benzoate (X_1), pyruvate (X_2), acetaldehyde (X_3) and formate (X_4). To ease later calculations these are numbered first, followed by the exchangeable metabolites, as described in Section 1.2.1.3. The numbering system is shown in Figure 2.2. It is assumed that the cofactor requirement is not limiting as it is balanced over the pathway. It is also assumed that secretion is not limiting. In addition, no mechanism for toxicity of either the substrate, the intermediate or the products is included in the model.

2.1.1 Theory

The next step in modelling is to express the reaction system in numerical terms. This is done by writing down the metabolic reaction matrix for the system, which contains the stoichiometric coefficients of all of the components in every reaction. (This is the same as the stoichiometric matrix described in Section 1.1.3.1.2.) In this matrix the column j represents the reaction j and we write in this column at row i:

$+\alpha$ if the reaction j produces α moles of metabolite i

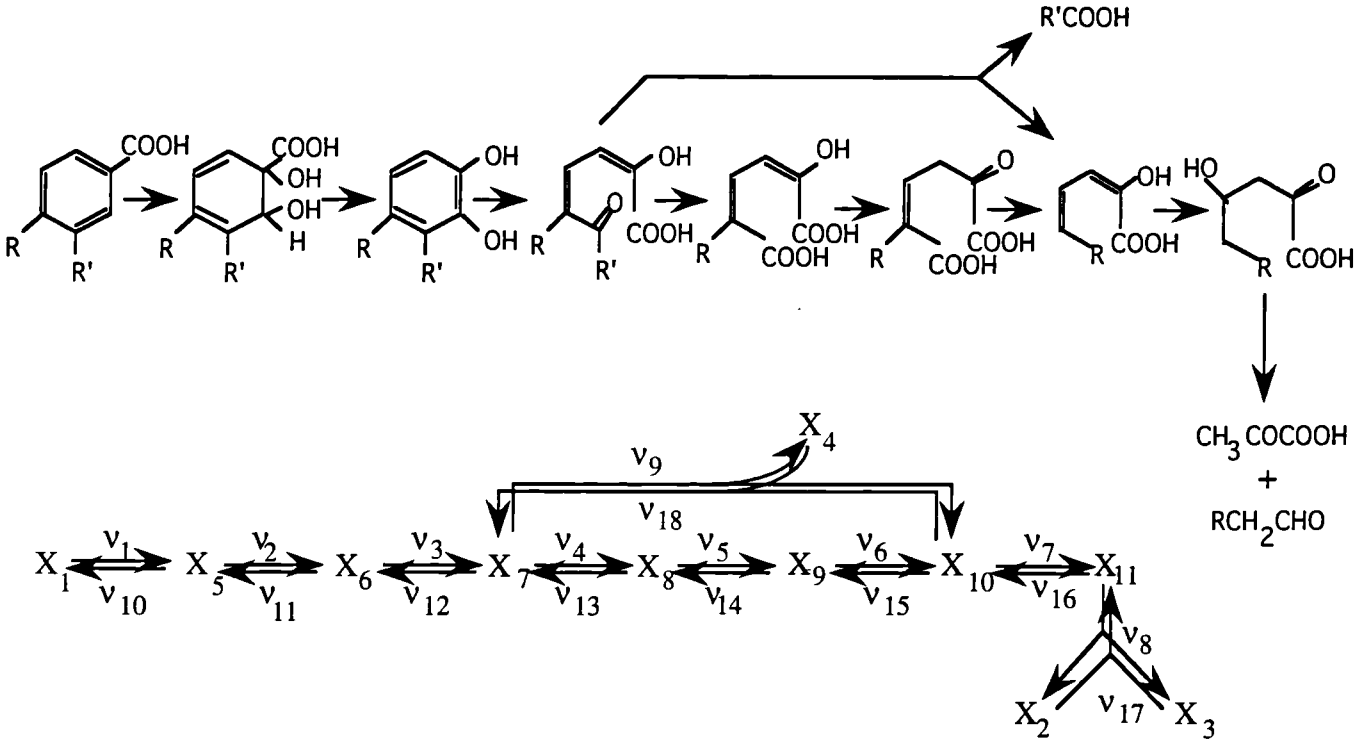


Figure 2.2: The TOL meta-cleavage pathway and its numerical representation.

	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	ν_9	ν_{10}	ν_{11}	ν_{12}	ν_{13}	ν_{14}	ν_{15}	ν_{16}	ν_{17}	ν_{18}	$\nu_A^{X_1}$	$\nu_A^{X_2}$	$\nu_A^{X_3}$	$\nu_A^{X_4}$
X_1	-1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	-1	0	0	0
X_2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	-1	0	-1	0	0
X_3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	-1	0	0	-1	0
X_4	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	-1
X_5	1	-1	0	0	0	0	0	0	0	0	-1	1	0	0	0	0	0	0	0	0	0	0
X_6	0	1	-1	0	0	0	0	0	0	0	0	-1	1	0	0	0	0	0	0	0	0	0
X_7	0	0	1	-1	0	0	0	0	-1	1	0	0	-1	1	0	0	0	0	0	0	0	0
X_8	0	0	0	1	-1	0	0	0	0	0	0	0	0	-1	1	0	0	0	0	0	0	0
X_9	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	-1	1	0	0	0	0	0	0
X_{10}	0	0	0	0	0	1	-1	0	1	-1	0	0	0	0	0	-1	1	0	0	0	0	0
X_{11}	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	-1	1	0	0	0	0

Figure 2.3: Stoichiometric matrix for the TOL degradation pathway.

$-\alpha$ if the reaction j consumes α moles of metabolite i

0 if the reaction j neither produces or consumes metabolite i

The structure of the system reaction matrix for a system with r reactions and m metabolites is shown in Figure 2.3 along with the metabolic reaction matrix **R** for the TOL degradation pathway.

The m metabolites can be divided into m' non-exchangeable compounds (those which remain inside the system) and $m - m'$ exchangeable compounds so that **R** is an m x $(r + m - m')$ matrix. The system rate vector v_{tot} has dimensions $(r + m - m')$ x 1 and we can write the set of conservation relations as

$$R \cdot v_{tot} = 0 \tag{75}$$

The metabolic reaction matrix **R** and the rate vector v_{tot} can be partitioned, separating the net conversion rates and the metabolic reaction rates. We write the m x 1 vector of net conversion rates as v_A:

$$N' \cdot v = v_A \tag{76}$$

where **N'** is m x r and v is r x 1. Considering only the non-exchangeable compounds we can write

$$N \cdot v = 0 \tag{77}$$

where **N** is m' x r.

Not all of these reactions are necessarily independent and we can extract from **N** a subset of its rows which is a basis of the whole set of its rows i.e. the rows of the basis set are independent and every row of **N** is a linear combination of the basis rows. Assuming that the first m_0 rows form the basis $\mathbf{N_R}$ (as it is always possible to renumber the rows) **N** can be decomposed as follows

$$N = K N_R \tag{78}$$

where the m' x m_0 matrix **K** has the form

$$K = \begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & 0 & 1 & 0 \\ 0 & \dots & \dots & \dots & 0 & 1 \\ --- & --- & --- & --- & --- & --- \\ & & & K_0 & & \end{bmatrix} \tag{79}$$

and $\mathbf{K_0}$ is a $(m' - m_0)$ x m_0 matrix.

The above decomposition illustrates one of the tests that must be made on a system before further analysis - that of independence. There are five such checks which should be made. These are independence, consistency, observability (Noorman *et al* 1991), existence of a steady state and stability of a steady state (Reder 1988).

The independency test can be stated as follows

$$rank(\boldsymbol{R}) = m \tag{80}$$

or

$$rank(\boldsymbol{N}) = m' \tag{81}$$

If rank(R) $< m$ then this indicates the presence of identical component relations. If rank(N) $< m'$ the presence of one or more dependent reactions is indicated. There are several ways of forming an independent set. Either the components or reactions responsible can be eliminated or the elements and reactions can be concentrated into an independent set. For example when considering the redox state of the cell we can replace NAD^+ and $NADH_2$ by the couple $NAD^+/NADH_2$ in systems where they are conserved and similarly for ATP, ADP and AMP which determine the energy balance within a cell.

The system must also be tested for consistency, that is, whether the mathematical relations are compatible. This involves finding the number of degrees of freedom of the system. This is the difference between the number of *a priori* unknown rates $(m - m' + r)$ and the number of independent relations specified by **R** (m_0) and gives the number of rates that must be calculated to allow the calculation of all of the others i.e.

$$\text{no. of degrees of freedom, } d_r = r - m' \tag{82}$$

If m' $\geq$ r then the system is either over-specified or exactly specified. In this case it is likely that a metabolic reaction is missing so that the system metabolic pathways need to be re-examined (Noorman *et al* 1991, Hofmeyr 1986).

The observability test ensures that the number of rates measured is sufficient for determination of the state of the system. The vector of measured rates v_A^m must have the same dimension as the number of degrees of freedom d_r. The metabolic reaction rates can be calculated by partitioning the matrix N' to separate the measured rates from the calculated rates:

$$\begin{aligned} (N')^m . v &= v_A^m \\ (N')^c . v &= v_A^c \end{aligned} \tag{83}$$

The criteria for observability is that

$$rank(N^m) = r \tag{84}$$

i.e. r linearly independent relations are required to determine the $v_1,\ldots, v_r$ reaction rates.

Metabolic control theories consider the perturbations of the system around a steady state and it is necessary to consider the existence and the nature of this steady state. To do this we need to calculate the Jacobian of the system $\mathbf{N}.v$. It can be shown that the system will reach a new steady state upon perturbation if the matrix $\mathbf{N_R}\nabla_x v\mathbf{K}$ is invertible (Reder 1988). $\nabla_x v$ is the r-row and m-column matrix of the partial derivatives of the rates with respect to the X metabolites. The steady state is stable if the real parts of all of the eigenvalues of $(\mathbf{N_R}\nabla_x v\mathbf{K})$ are negative (Reder 1986).

Having simulated and examined this steady state we need a linear way of modelling the action of the system around it, with respect to changes in the system parameters. Because of the non-linear form of most enzymatic rate laws metabolic models are not numerically tractable if we wish to perform more sophisticated analyses especially for larger systems. By transforming into logarithmic coordinates it is possible to obtain a linear expression for metabolic systems.

2.2 Simulation

2.2.1 Theory

To simulate a metabolic system, the power-law approximation which forms the basis of biochemical systems theory as described in Section 1.2.2 will be used. The power-law formalism is analogous to a linear model, being based on the first two terms of the Taylor series, but in logarithmic space. The use of log coordinates was suggested by the fact that the rational functions which provide a good representation for rate laws *in vitro* can be approximated over a wide range by a straight line on a log-log plot [Savageau and Voit, 1982; Savageau *et al,* 1987a]. A rate expression similar to the law of mass action for chemical reactions is obtained by writing the Taylor series for the rate of reaction v_i for each of the reactions in the pathway about an operating point and transforming back into Cartesian coordinates so that the rate can be expressed as a product of power-law functions [Voit and Savageau, 1982; Savageau *et al* 1987a].

$$v_i(X_1,\ldots,X_n) = \alpha_i \prod_{j=1}^{n} X_j^{g_{ij}} \tag{85}$$

where

$$\alpha_i = v_{iO} \prod_{j=1}^{n} X_{jO}^{-g_{ij}} \tag{86}$$

$$g_{ij} = \partial(\ln v_{iO})/\partial(\ln X_j) = (\partial v_{iO}/\partial X_j)(X_{jO}/v_{iO}) \tag{87}$$

Using this formalism to model a metabolic system we apply the S-system variant (see Section 1.2.2.2.1) [Savageau, 1991]. The essential feature of this variant is in the aggregation of those fluxes which tend to increase a given substrate into one net rate law and those which tend to decrease into another [Sorribas and Savageau, 1989c]. The power-law formalism explicitly represents each of these net rate laws so that for each substrate (or pool of substrates) X_i we can write the equation

Here the $ß_i$ and h_{ij} are respectively the kinetic orders and rate constants for the rate laws

$$dX_i/dt = V_i - V_{-i} = \alpha_i \prod_{j=1}^{n} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n} X_j^{h_{ij}} \qquad i = 1,...,n. \tag{88}$$

decreasing X_i. There is one equation for each of the m' dependent variables and the product is over n where n-m' is the number of independent variables.

As described in Section 1.2.1.2.2 Eqn.s 33-38, the time derivatives can be set to zero for steady state and the system can be written in matrix notation and partitioned to solve for an explicit solution.

From this solution can be calculated the sensitivities of the system $S(X_i, X_j)$, $S(X_i, \alpha_j)$ and $S(X_i, g)$ (see Eqn.s 39-42 for definitions) to determine the rate-controlling parameters. These sensitivities indicate the presence of any bottlenecks in the process and what variables must be altered to remove them in that they relate the fractional change in any dependent variable to the fractional change in the independent variable responsible for the change. These sensitivities are equivalent to flux and concentration control coefficients in metabolic control theory.

2.2.2 A program for the Simulation of Metabolic Systems, **MetaSim**:

1. Simulation of system using kinetic data.

The aim of this program is to simulate the steady-state action of the system using kinetic rate laws and parameters for the reaction rates. The rate of change of the concentration of each component is described by rate laws for each reaction associated with it. A quasi-steady state is then assumed, which sets the rate of change of the intermediate concentrations equal to zero, and the resulting set of relations are solved for the steady state intermediate concentrations. One and two substrate Michaelis-Menten kinetics are assumed. The form of these laws is as follows: for the one substrate reaction:

$$r_X = v = \frac{k_1 [E] [X]}{K_m + [X]} \tag{89}$$

For a two substrate reaction

The metabolites in the system are grouped into m' non-exchangeable components -

$$-\frac{dX_1}{dt} = -\frac{dX_2}{dt} = v = \frac{k_1[E]}{\left[1 + \frac{K_1}{[X_1]} + \frac{K_2}{[X_2]} + \frac{(K_2 K_{12} + K_1 K_{21})}{2[X_1][X_2]}\right]} \quad (90)$$

those which are retained within the system and m-m' exchangeable components - those which are exchanged with the environment, and are numbered $X_1,\ldots,X_{m'},X_{m'+1},\ldots,X_m$. The system reactions are numbered $r_1,\ldots,r_r$. The reaction scheme is described numerically by a stoichiometric matrix. In this matrix the column j represents the reaction j and we write in this column at row i :

$+\alpha$	if reaction j produces α moles of metabolite i
$-\alpha$	if reaction j consumes α moles of metabolite i
0	if reaction j neither produces nor consumes metabolite i

The form of the stoichiometric matrix can be seen in Figure 2.3 which illustrates the stoichiometric matrix for the TOL pathway. The use of a rigorous nomenclature ensures that systems of any complexity can be dealt with once the above numbering scheme is used.

The file **InputData.m** contains all of the input information for the subsequent modules. The kinetic data is contained in a matrix **k**, where $k_{i,j}$ is the jth kinetic parameter of the ith reaction. The order of the kinetic parameters is as follows:

For a one substrate reaction:	k_1	K_m			
For a two substrate reaction	k_1	K_1	K_2	K_{12}	K_{21}

[Note: While the method of ordering and numbering of the reactions and metabolites is fixed, the choice of kinetics and kinetic parameter numbering is not. The kinetic expressions appear in the functions **kinexp** in the module **ModelSolve.m.** Any form of kinetic expression can be used here and the kinetic parameters should be numbered to correspond to these functions.]

The enzyme concentrations are entered in the array **enzyme**. The exchangeable metabolite concentrations are contained in the array **xdep**, and initial guesses for the non-exchangeable concentrations in the array **xindep**.

The file **ModelSolve.m** solves for the steady state values of the non-exchangeable metabolites. It is comprised of three functions: **RateExp**, **EqnSet** and **Findroot.**

RateExp contains the kinetic expressions **kinexp** and uses the stoichiometric matrix **r** and the matrix of kinetic parameters **k** to return the array of rate expressions for each reaction in the system (**v**).

EqnSet displays the full set of equations describing the rate of change of each metabolite in the system. It is by setting the rate of change of the non-exchangable metabolites equal to zero that the steady state values are found.

Findroot solves the set of equations obtained from **EqnSet** using input values for the exchangable metabolite and enzyme concentrations and initial guesses for the non-exchangable metabolite concentrations (**xdep, enzyme** and **xindep**, respectively). This set of simultaneous equations is solved using the Mathematica subroutine **FindRoot**. **FindRoot** searches for a numerical solution to a system of equations using either Newtons method or a variant of the secant method. The steady state values are returned in **output.**

findroot returns the non-exchangable metabolite concentrations for one set of kinetic parameters and enzyme and exchangable metabolite concentrations but what is more often required are the concentration profiles obtained when one of the input parameters is varied. The function **FindRoots** plots the concentration profiles obtained when either the input concentrations, the enzyme concentrations or the kinetic parameters are varied. To investigate the effects of changing an input parameter, it is sufficient to enter the range (and step size) over which it is to be varied in place of the usual single value. the change in the data entry format allows the subroutine to recognise which parameter is to be varied and **Findroot** is applied at each value. The concentration profiles are returned in plotted in a file called **plot1**.

While this approach works well for small and medium sized metabolic systems, for large or complex systems the matrices involved become difficult to solve and it is important to have close initial guesses. For this reason such solution techniques must be applied with care to large systems.

The purpose of the module **PowerLaw.m** is to calculate simulate the metabolic system on the basis of a power-law formalism. It has been shown that this type of model is suitable for describing the characteristics of metabolic systems and it can be posed in a linear form which greatly simplifies the numerical analysis.

The first function of the module, **KinParam**, uses the array of rate expressions v from **RateExp** and the steady state values in output to calculate the power law parameters α, **ß, g** and **h.**

Recall that the system can be described by an expression of the form

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^{m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{m} X_j^{h_{ij}} \qquad i = 1,\ldots,m_0. \tag{91}$$

SSCalc uses these parameters to solve the matrix equation

$$\mathbf{A}y = b \tag{92}$$

where

$$\begin{aligned} y_i &= \ln X_i \\ b_i &= \ln(\beta_i / \alpha_i) \\ a_{ij} &= g_{ij} - h_{ij} \end{aligned} \tag{93}$$

PL performs the same function here as **FindRoots** in **ModelSolve.m**, in that it applies **SSCalc** repeatedly over a range of one of the input parameters to return profiles of the concentrations and rates in files named **xplot** and **vplot.**

Full listing and documentation of this software is given in Appendix D.

2.3 Optimization

2.3.0 Introduction

While optimization techniques have been used as a tool in chemical engineering process design for many years, their applications in the bioprocess industries have so far been limited. There are several reasons for this. The primary reason is that the most common solution techniques have been developed for well-defined linear problems and there are few bioprocess which fall into this category. However as solution techniques improve and as understanding of bioprocess improves, the number of cases in which optimization techniques can provide useful insights and data is increasing.

The goal of optimization is to find the values of the variables in the process that yield the best value of the performance criterion. In the formulation of an optimization problem it is necessary to identify firstly *the objective function*, which represents the performance criterion which is to be optimized and this must be expressed in terms of the key variables of the process being analysed, and secondly *the process model*, which describes the interrelation of the key variables and provides the constraints for the solution.

By describing a process in mathematical terms a variety of computational techniques can be used to identify the best solution from the set of possible solutions. However these must be used intelligently if the results are to be meaningful. This requires insight as to the appropriate objective function to be used and good judgement in the interpretation of the results.

The essential features of an optimization problem are: one or more objective functions to be optimized, equality constraints and inequality constraints. The latter two arise from the model describing the behaviour of the system. An optimal solution is one which satisfies these constraints (i.e. is a feasible solution) and which also supplies the optimal value to the objective function). In most cases the optimal solution is a unique one. If there are no residual degrees of freedom among the constraints then there exists only one solution and there is no need to search for an optimum.

When solving an optimization problem there are several steps which should be

followed.

1) List all of the variables of importance in the system

2) Identify the objective function and define it in terms of the system variables.

3) Develop the expressions which will form the system model. Identify the independent and dependent variables to determine the number of degrees of freedom.

4) Apply a suitable optimization technique to the mathematical model of the system.

5) Examine the sensitivity of the model, that is, the effect on the result of changes in the input parameters.

In this section we describe how this approach can be applied to a metabolic system in conjunction with the modelling techniques described previously.

2.3.1 Theory

The previous sections have described how a metabolic system can be described in power-law terms so that it becomes linear in logarithmic coordinates. Having expressed the problem in a linear form it is possible to apply standard optimization techniques. In the case of metabolic pathways the objective function will usually be either a flux or a metabolite. That is, we will wish to maximize or minimize either the rate at which a metabolite is produced or its level in the cell. The constraints on this system are imposed by the pseudo-steady state assumption applied to the non-exchangeable elements of the system. The problem is a constrained linear optimization problem and of the many linear programming techniques available the chosen solution method is by a NAG library subroutine E04MBF.

A general way of expressing a linear programming problem is

$$\underset{x}{minimize} \quad f(x) \triangleq c^T x \tag{94}$$

subject to

$$a_i^T x = b_i \qquad i \in E \tag{95}$$

$$a_i^T x \geq b_i \qquad i \in I \tag{96}$$

Eqn.(95) is the set of equality constraints and Eqn.(96) is the set of inequality constraints. When considering a metabolic pathway, the objective function will be either a metabolite or a flux i.e.

$$f(x) = \ln x_{opt} \tag{97}$$

or

$$f(x) = \ln V_{opt} = \ln \alpha_{opt} + \sum_{j=1}^{n} g^{optj} \ln X_j \tag{98}$$

subject to the equality constraint given by Eqn.(95) and the inequality constraints $x_i > 0$ i=1,...,n. The equality constraints arise from the linear description of the problem, Ax=b, and the inequality constraints form the physical limitations of the system.

2.3.2 A Package for Optimizing Metabolic Pathways MetaOpt

This software is based upon the same subroutines as that for simulating metabolic pathways, **MetaSim**, and makes use of the power law form employed by that package to express the system in linear form and apply linear optimization techniques. In this way, the optimum set of parameters for a particular objective function can be determined. Full listing and documentation is given in Appendix D.

The parameters to be optimized are the intermediate and the enzyme concentrations. These are subject to a set of inequality and equality constraints. The inequality constraints are in the form of upper and lower bounds and these ensure that the solution is physically reasonable. The lower bound specifies that the concentrations should not be less than zero. The upper bound will depend on a range of factors determined by the particular system: upper bounds on intermediate concentrations may be specified by their levels of toxicity to the cell while upper bounds on enzyme concentrations will depend on the maximum expression level of the enzymes in the organism under process conditions.

The software which performs this optimization is called **MetaOpt**. It is similar to the package **MetaSim** described in Section 2.2.2 in that it calculates the steady state operating point of a metabolic system from its kinetic parameters and then determines the power law parameters for the system from this operating point. Instead of simulating the system about this operating point, it uses the linear form of the power law expression in logarithmic coordinates as the equality constraints for the optimization routine.

The package **MetaOpt** contains the functions **Findroot, RateExp, KinParam,** and **SSCalc** which are as described in Section 2.2.2 and the additional function **Opt**. The arguements required by **Opt** are the same as those required by **PL (k, aa, int, ext, enzime)** with the addition of the objective function **c**. The upper and lower bounds specified in the program as upper and lower can be specified or the default values of upper = 100. and lower = -10. can be used.

The optimization routine used here is the NAG library subroutine **e04mbf** (see Appendix D for details and parameter listing) and was chosen because of its robustness and ease of use. **MetaOpt** writes the objective function, the system matrix A, the upper and lower bounds and the starting point (which is the operating point) to a data file **modsol.r**. The Fortran subroutine is then compiled and executed from within **MetaOpt** and the results from the output data file **optsol.r** read in and plotted.

2.4 Fermentation Data

2.4.0 Introduction

The previous sections have described the simulation and optimization of a metabolic system using kinetic data. This approach is useful in that it can provide initial data on a system from *in vitro* data on the individual reactions. It may also give insight into the behaviour of reactions which for practical reasons cannot be monitored *in vivo*. However, results obtained from this method may be misleading if the kinetic experiments have been carried out under very different conditions from those expected

in vivo or if elements of the system behave significantly differently *in vitro* than *in vivo*. For these reasons, it is necessary to validate results acquired from kinetic data with data from the system *in vivo*. There are several ways of achieving this, in general based on the variation of levels of enzyme and intermediate expression.
In this section the analysis of *in vivo* data for the determination of control coefficients is described.

2.4.1 Theory

The dynamic approach uses measurements of the concentration profiles of metabolites following a perturbation to the system to measure the flux control coefficients. Because the transient metabolite concentrations do not vary freely but are constrained [Reder, 1986] the relationships between these concentrations can be used to evaluate the flux control coefficients [Delgado and Liao, 1992].
The transient metabolite concentrations can be correlated with time by the following expression

$$\sum_{i=1}^{n+1} \alpha_i (x_i(t) - x_i(0)) = t \qquad (99)$$

where $x_i(t)$ are the metabolite concentrations at time t for n metabolites and α_i are the correlation coefficients which are evaluated from linear regression. The flux control coefficients can then be calculated from

$$[C^J_{E_1}\, C^J_{E_2},\ldots,C^J_{E_n}] = [\alpha_1\, \alpha_2,\ldots,\alpha_n]\, A\, J \qquad (100)$$

Here, A is the stoichiometric matrix shown in Figure 2.3, and J is the steady-state flux. This is derived from a linearization of the kinetics as follows:

$$r_i = \sum_{j=1}^{L-1} a_{ij} X_J + b_i \qquad (101)$$

or generally:

$$r_i = \sum_{j=0}^{L} a_{ij} Y_j + b_i \tag{102}$$

where $Y_0=s$, $Y_j=X_j$, $j=1,2,\ldots,L-1$, $Y_L=p$

$$\epsilon_{ji} = \frac{Y_j}{r_i} a_{ij} \quad \left(= \frac{\partial r_i}{\partial Y_j} \frac{Y_j}{r_i} \right) \tag{103}$$

$$0 = \sum_{i=1}^{L} C_i^r a_{ij} \frac{X_j}{r_i} \tag{104}$$

for fixed s and p.

$$0 = \sum_{i=1}^{L} C_i^r a_{ij} \tag{105}$$

since X_j are independent of i and $r_i = r$ at steady s.

$$0 = \sum_{i=1}^{L} C_i^r a_{ij} \Delta X_j \tag{106}$$

with multiplication by a constant.

$$0 = \sum_j \sum_i C_i^r a_{ij} \Delta X_j = \sum_i C_i^r \sum_j a_{ij} \Delta X_j = \sum_i C_i^r r_i \tag{107}$$

where $\Delta X_j = X_j(t) - X_j(t=0)$ which is the transient response. Hence C_i^r can be obtained by least squares from measurement of the interval rates (or ΔX_j) at an arbitrary number of time values $t_k > 0$.

2.4.2 Description of *In Vivo* Data Program, **InVivo**

The most important part of this program is the routine for linear regression supplied by the NAG subroutine **G02DAF**. This routine performs a general multiple linear regression and computes parameter estimates, standard errors, residuals and influence statistics (see Appendix D for a description and listing of the subroutine parameters). The input data required for this program module is an array of time-varying metabolite concentrations, the stoichiometric matrix, and the system fluxes. The concentration data is read into the file **res.d** and the FORTRAN program **residual.f** is compiled and the object file **a.out** executed by **InVivo**. The calculated parameters, α, are then read back into Mathematica and the control coefficients calculated using the flux array **J** and the stoichiometric matrix **R**.

2.5 Sensitivity Analysis

2.5.0 Introduction

An essential consideration of any system model is how it responds to small perturbations in its input data. The measure of this response is termed the sensitivity of the system. It is essential that the sensitivity of a system be such that it can cope with experimental errors and noise. For linear systems it can be shown that the change in solution upon perturbation of initial data can be estimated by a condition number, k. For non-linear systems, it is possible to define upper and lower bounds which reduce to the condition number k under linear approximation.

2.5.1 Theory

2.5.1.1 Linear Systems

Simple matrix theory can be used to investigate the linear system

$$Ax = b \tag{108}$$

with nonsingular A $\in$ L($\mathbb{R}^n$), and determine the effect of perturbations in A and b on x. Let the perturbation of this system be a matrix B $\in$ L($\mathbb{R}^n$) which is close to A in that

$$\|A^{-1}\| \, \|B-A\| < 1 \tag{109}$$

Then we can show that B is also nonsingular and that for b $\neq$ 0, the solutions $x^* = A^{-1}b$, $y^*=B^{-1}c$ satisfy the estimate

$$\frac{\|x^*-y^*\|}{\|x^*\|} \leq \frac{\kappa(A)}{1 - \kappa(A)\|B-A\|/\|A\|}\left[\frac{\|B-A\|}{\|A\|} + \frac{\|b-c\|}{\|b\|}\right] \tag{110}$$

where

$$\kappa(A) = \|A\| \, \|A^{-1}\| \tag{111}$$

is the condition number of A, i.e. the relative error in x can be κ times the relative error in A and b and in this way κ quantifies the sensitivity of the linear system. It should be noted that the condition number is norm-dependent. There are several matrix norms which are generally used. The most common of these are the Frobenius norm (F-norm)

$$\|A\|_F = \left[\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2\right]^{1/2} \tag{112}$$

and the p-norms

$$\|A\|_p = \sup_{x\neq 0} \frac{\|Ax\|_p}{\|x\|_p} \tag{113}$$

Any two condition numbers $\kappa_\alpha(A)$ and $\kappa_\beta(A)$ are equivalent in that there exists constants c_1 and c_2 for which

$$c_1\kappa_\alpha(A) \leq \kappa_1(A) \leq c_2\kappa_\alpha(A) \tag{114}$$

2.5.1.2 Non-Linear Systems

For non-linear systems κ does not exist and it is necessary to define other quantities which describe the sensitivity of the system. These quantities are the greatest lower bound and the least upper bound or Lipschitz norm.

For X and Y which are real normed linear spaces and L(X, Y) the space of bounded linear operators from X into Y. For any mapping F: D $\subset$ X $\Rightarrow$ Y and closed subset C$\subset$D, we introduce the greatest lower bound

$$\mu(F, C) = sup\{ t \in [0, \infty);\ \|Fx - Fy\| \geq t\|x - y\| \quad \forall\ x,y \in C \} \tag{115}$$

and the least upper bound or Lipschitz norm

$$\nu(F, C) = inf\{t \in [0, \infty] ;\ \|Fx - Fy\| \leq t\|x - y\| \quad \forall\ x,y \in C\} \tag{116}$$

This suggests the following definition of the condition number for complex systems

$$\kappa(F, C) = \begin{cases} \nu(F, C)/\mu(F, C) & \text{if } 0 < \mu(F C), \nu(F, C) < \infty \\ \infty & \text{otherwise} \end{cases} \tag{117}$$

(Rheinboldt, 1976)

2.6 Results

2.6.0 Introduction

Two systems will be examined. These illustrate the principle of top-down analysis in that each is an increasingly detailed version of the same system. The simplifications are based upon the data which can be collected. In examining a system in such a manner we are trying to match what should be measured with what can be measured. By using the fermentation data as our simplification criteria we can obtain initial results which

can then be used to determine what are the important areas of the system which should be further investigated. In this way, this type of analysis can be used to direct the analysis required for pathway engineering.

2.6.1 The two-step system

This system is illustrated in Figure 2.4. It shows the TOL meta-cleavage pathway simplified to two steps, involving the compounds benzoate, catechol, pyruvate and acetaldehyde. This simplification was initially chosen because of the ease of availability of these compounds for use as standards and because methods for analysis of these compounds were easy to develop.

Following the procedure for establishing and validating a system model as outlined above:

The stoichiometric reaction matrix is

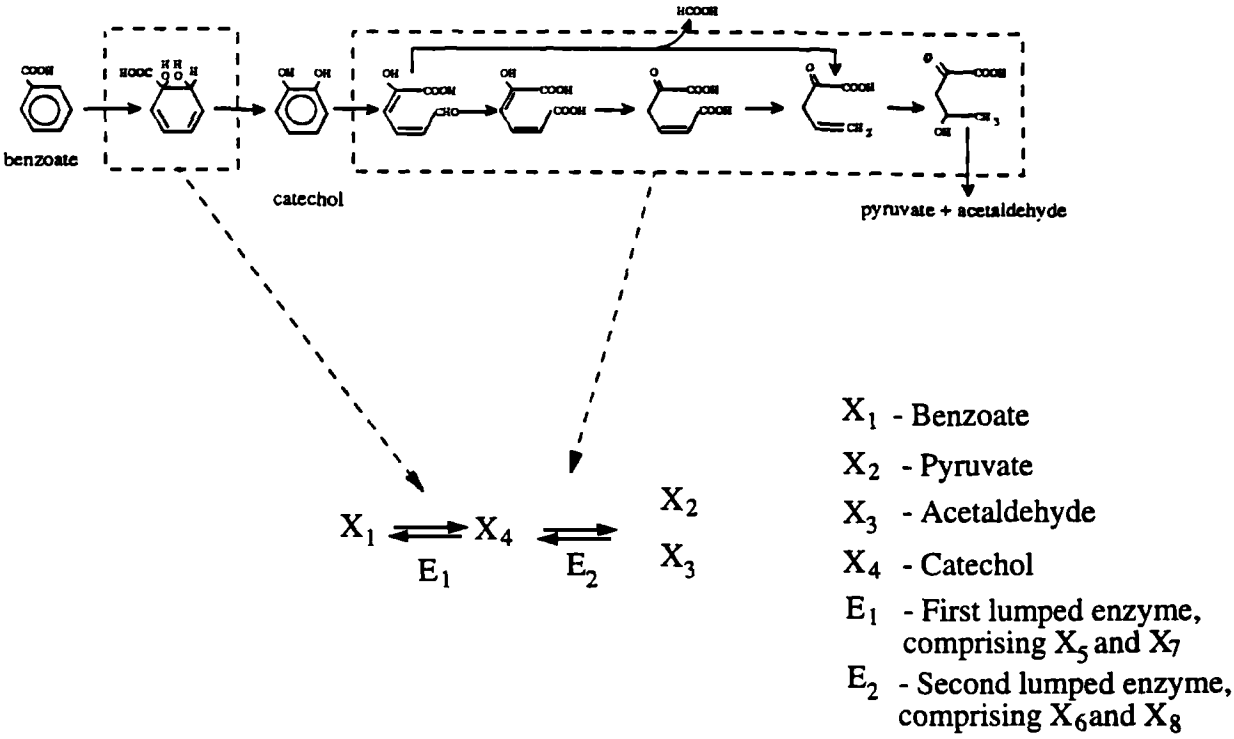
$$\boldsymbol{R} = \begin{bmatrix} -1, & 0, & 1, & 0, & -1, & 0, & 0 \\ 0, & 1, & 0, & -1, & 0, & 1, & 0 \\ 0, & 1, & -1, & 0, & 0, & 0, & 1 \\ 1, & -1, & -1, & 1, & 0, & 0, & 0 \end{bmatrix} \tag{118}$$

We can partition this matrix, separating the net conversion rates and the metabolic reaction rates to yield the m x r matrix N'

$$\boldsymbol{N'} = \begin{bmatrix} -1, & 0, & 1, & 0 \\ 0, & 1, & 0, & -1 \\ 0, & 1, & 0, & -1 \\ 1, & -1, & -1, & 1 \end{bmatrix} \tag{119}$$

and from this we can isolate the simplified reaction matrix, N involving only the non-exchanged compound, X_4

$$\boldsymbol{N} = [1, \ -1, \ -1, \ 1] \tag{120}$$



ıre 2.4: Two-step lumped system. Note: the production of formate is small so it has been omitted in this

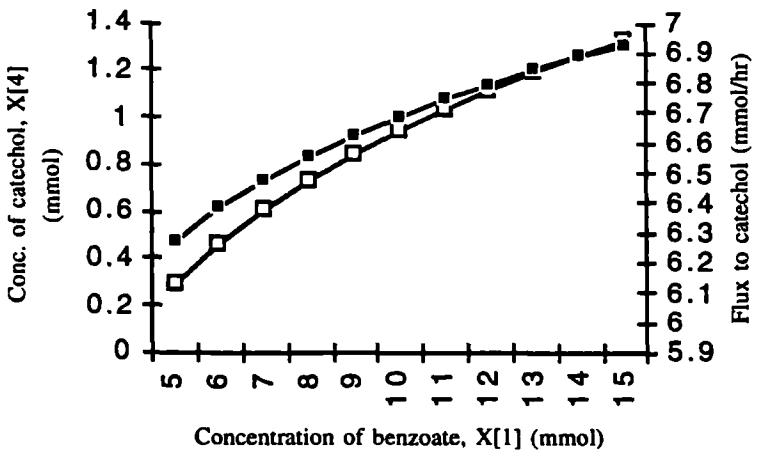


Figure 2.5 Predicted variation of catechol levels (❑) and pathway flux (■) with increasing benzoate concentration for lumped system.

Because of the simplicity of this system, the results of some of the validation tests are obvious, but it is good practice to carry them out routinely. It is obvious that the system is a linearly independent one as it comprises only one equation. This can be tested by calculating the rank of the matrices **R** and **N** (see Appendix A for a definition of rank and some other concepts in matrix algebra).

$$\text{rank}(R) = 4 = m \text{ as required}$$
$$\text{rank}(N) = 1 = m' \text{ as required}$$

The number of degrees of freedom = r-m' = 3. This indicates that the system is consistent and that it is necessary to measure three of the system components or rates to determine the state of the system.

Partitioning the matrix N' to separate the measured rates from the calculated rates and applying the test for observability, we find $\text{rank}((N')^m) = 4 = r$ so that 4 linearly independent reaction relations are required to determine the $v_1,...v_4$ reaction rates if the criterion of observability is to be fulfilled.Having established the independence, consistency and observability of the system it is now necessary to calculate the steady state of the system and to investigate the existence and stability of the steady state resulting from a perturbation of this system. To do this we use Michaelis-Menten kinetic parameters obtained from the literature (see Table 2.1) to calculate the pseudo-steady state value of X_4. This value is X_4 = 2.57 mmol.

Table 2.1: Kinetic parameters for lumped system.

Reaction	k_1 (s^{-1})	K_m (mM)	K_2	K_{12}	K_{21}
v_1	100	15			
v_2	70	8			
v_3	50	3			
v_4	1.5	4	6	3	9

To simulate the system about this operating point we use a power-law expression where the coefficients and exponents are calculated from the kinetic parameters using the subroutine **KinParam** and for this system the power-law expression for the change in

the pool of catechol is:

$dX_4/dt = -22.09\ x[4]^{0.63}x[6]^{0.42}x[7]^{0.57} + 10.08\ x[1]^{0.59}x[2]^{0.002}x[3]^{0.001}x[5]^{0.99}x[8]^{0.002}$ In

This expression, the rate of change of the pool of catechol (X_4) is dependent on the concentration of the lumped enzyme catalysing the first forwards reaction x[5] to the power of almost 1 and the concentration of the lumped enzyme catalysing the first backwards reaction (x[7]) to the power of 0.57. This situation would be unlikely in mechanistic terms if E_1 were an actual enzyme and would warrent a review of the model assumptions or the kinetic data but is not unfeasible when the enzyme is a result of the lumping together of several enzymic steps.

Using this expression the profiles for the non-exchangeable intermediates and fluxes which occur as a result of changes in the exchangeable parameters can be calculated. Figure 2.5 shows how X_4 and V_4 change with varying X_1 (input benzoate concentration). In this example the input benzoate concentration is varied from 5 mmol to 15 mmol (the operating point about which the system is modelled is at 10 mmol). Over this range it can be seen that the catechol concentration increases from just over 6 mmol to 7 mmol. The flux to catechol is also shown and this increases from approximately 0.5 mmol/hr to 1.3 mmol/hr over the range examined. The approximation is exact with respect to the Michaelis-Menten approximation at the operating point and is accurate to over 98% even at the edges of the range (data not shown for the lumped system, see Figure 2.13 for results for the full system). Similar profiles can be obtained for any of the exchangeable parameters i.e. the input, output and enzyme concentrations.

The flux and concentration control coefficients are found using the coefficients g_{ij} and h_{ij} and the matrices **L** and **M** from Eqn.(37). The concentration control coefficients are found from the expression

$$\frac{\partial \ln X_i}{\partial \ln X_j} = \frac{\partial y_i}{\partial y_j} = L_{ij} \qquad (1)$$

and the flux control coefficients from

$$\frac{\partial \ln V_i}{\partial \ln X_j} = g_{ij} L_{ij} \qquad (2)$$

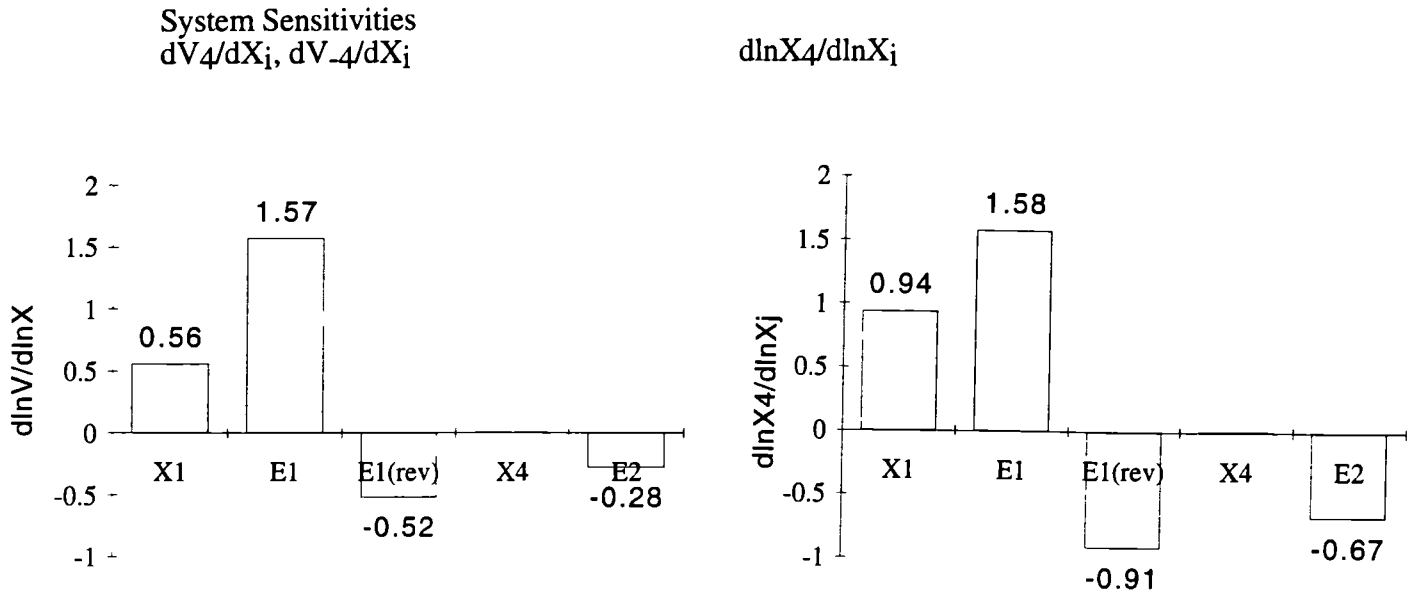


Figure 2.6: System sensitivities for the two-step system shown in Figure 2.4. These show clearly that the bulk of the control resides in E_1 with X_1 also playing an important part.

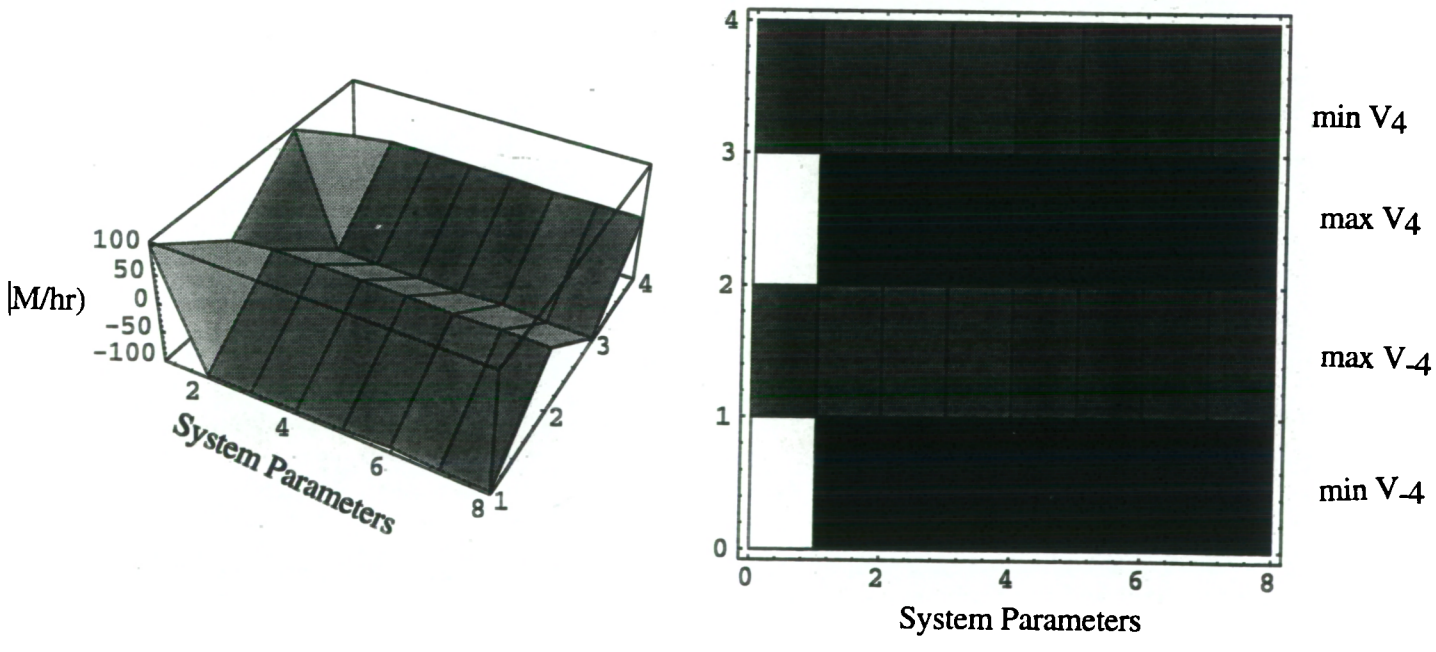
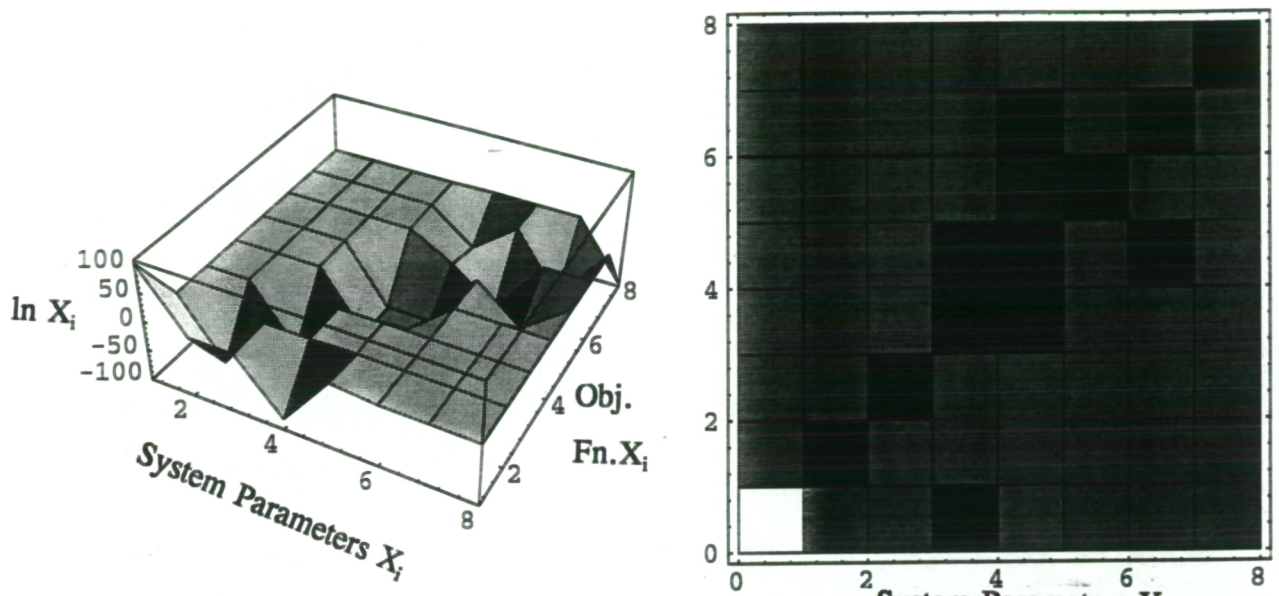


Figure 2.7: Requires system parameters for the maximization and minimization of $V_{+/-4}$. As expected, the required conditions for the maximization of V_4 are identical to those for the minimization of V_{-4}

(a) Minimization of X_i (c_i=1.0):



(b) Maximization of X_i (c_i=-1.0):

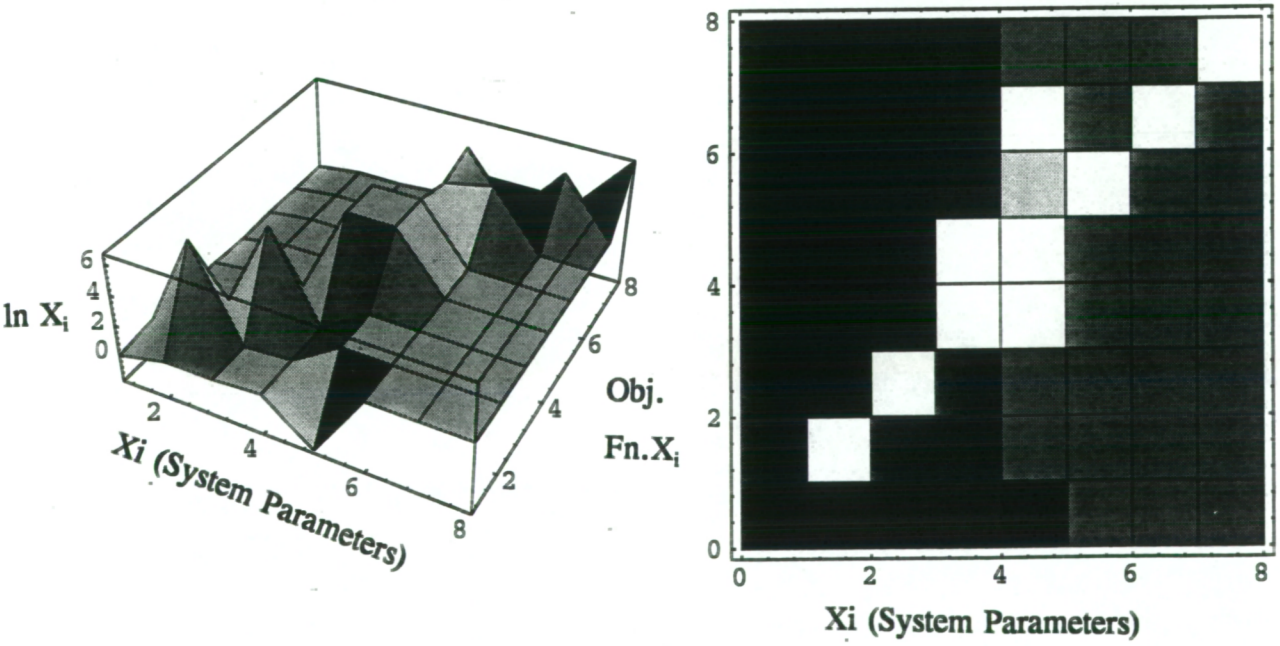


Figure 2.8: These 3-D and density plots show the levels of the system parameters required for the optimization of each of the system concentrations.

These coefficients are shown in Figure 2.6 which shows which components of the system have the greatest effect on the level of the intermediate X_4 (catechol), and on the fluxes which go to increase and decrease its concentration, V_4 and V_{-4}. The results are what would be expected from examination of the profiles in Figure 2.5 and also from the physiology of the system. The greatest controlling element in the system is the first lumped enzyme (comprising the effects of the first two enzymes in the pathway). This has a flux control coefficient of 1.57 and a concentration control coefficient of 1.58. Increasing the concentration of the second lumped enzyme in the pathway has a negative effect on both the flux to catechol and the concentration of catechol, as can be seen from the flux control coefficient of -0.52 and the concentration control coefficient of -0.91.

The linear form of the power-law expression can be used to optimize the system using standard linear optimization techniques. A very large amount of data is generated by this technique, as a full set of optimal system parameters is generated for each objective function tested. Because of this, the most convenient way of displaying the data is by a density plot or by a 3-D plot. Figure 2.7a shows the optimal state of the system values when each of the intermediate and enzyme concentrations is minimized in turn and Figure 2.7b shows the optima achieved when these concentrations are maximized. Figures 2.8a and 2.9b show the optima achieved when the system fluxes are minimized and maximized, respectively. Examining the predicted system parameters for the maximization or minimization of the system concentrations, it can be seen that they correspond with the system control coefficients. For example, to maximize or minimize the catechol concentration, X_4, it is necessary to maximize or minimize the concentration of the first enzyme in the system which reflects its high control coefficients, which is a predictable and physically reasonable result.

For flux optimization, it is the benzoate concentration which has the greatest effect. Because A is a 1x1 matrix the condition number κ is automatically 1 and the system is well-conditioned. This indicates that small errors in the input data will not cause a great error in the results.

2.6.2 The Full Pathway

This is the full pathway as shown in Figure 2.2. Applying the above tests to the TOL meta-cleavage pathway shown in Figure 2.2 and described by the reaction matrix **R**, above the following results were obtained:

In this set *rank* (**N**) = 7 = m'. This indicates that there are no dependent reactions in the set so that the reaction set is a basis and $\mathbf{N} = \mathbf{N_R}$ ($\mathbf{K} = \mathbf{I}$) and $m_0 = m'$.

The number of degrees of freedom in the system is $d_r = 18 - 7 = 11$ and the system is conceptually realistic.

Table 2.2: Kinetic parameters for the TOL meta-cleavage pathway.

	v_{max}	E	k_1	K_1	K_{12}	K_{21}
v_1	1600	210	7.62	75	-	-
v_2	12	210	0.05	62	-	-
v_3	1400	200	7	200(a)	-	-
v_4	25	195	0.13	89	-	-
v_5	128	180	0.71	30(b)	-	-
v_6	105	185	0.57	55	-	-
v_7	120	210	0.57	2(c)	-	-
v_8	5	210	0.02	11	-	-
v_9	6000(d)	100	600	20(d)	-	-
v_{10}	100	100	1	18	-	-
v_{11}	51(d)	195	0.26	15(d)	-	-
v_{12}	6	195	0.03	23	-	-
v_{13}	120(d)	230	0.52	30(d)	-	-
v_{14}	25	230	0.11	21	-	-
v_{15}	100	128	0.78	53	-	-
v_{16}	95	128	0.74	34	47	32
v_{17}	80(e)	250	0.32	100(e)	-	-
v_{18}	65	250	0.26	74	49	62

The check for observability is that the number of rates which must be measured to completely determine the system is equal to the number of degrees of freedom and in

this case the number is 11. The rank of the matrix $\mathbf{N}^m$ is 18 which is sufficient for observability.

The steady state which will be examined is simulated by choosing realistic kinetic parameters and metabolite concentrations of the same order of magnitude as those obtained from the literature (see Table 2.2). Michaelis-Menten kinetics are assumed to be valid and the simulated steady state is determined by solving the set of simultaneous rate equations to obtain the results in Table 2.3.

To show the existence of a single steady state upon perturbation the Jacobian of the system $\mathbf{N}.{}_x\mathbf{v}$ is calculated and shown to be invertible. The eigenvalues of this matrix are listed in Table 2.4. These are all negative, indicating the stability of the steady state.

To simulate the TOL meta-cleavage pathway using biochemical systems theory, the pathway is written as shown in Figure 2.2(b) so that each of compounds affecting the system is denoted by X_i. X_1 to X_4 are the exchangeable metabolites, X_5 to X_{11} are the non-exchangeable metabolites and X_{12} to X_{29} are the enzymes of the pathway. Using Eqn.s (30) and (31) the parameters α_i, β_i, g_{ij} and h_{ij} can be calculated from the kinetic parameters in Table 2.2 to give a power-law model for the system given in Table 2.3.

Table 2.3: Power-law approximation for the TOL pathway

$$dX_5/dt = -1.797\ x_5^{0.109}x_{13}^{0.889}x_{22}^{0.11} + 0.192\ x_1^{0.713}x_6^{0.001}x_{12}^{0.999}x_{23}^{0.001}$$

$$dX_6/dt = -14.719\ x_6^{0.959}x_{14}^{0.999}x_{23}^{0.001} + 1.832\ x_5^{0.030}x_7^{0.00003}x_{13}^{0.999}x_{24}^{0.00003}$$

$$dX_7/dt = -45.418\ x_7^{0.975}x_{15}^{0.999}x_{20}^{0.0003}x_{24}^{0.00003} + 9.22\ x_4^{0.034}x_6^{0.831}x_8^{0.055}x_{10}^{0.037}x_{14}^{0.866}x_{21}^{0.052}x_{25}^{0.080}$$

$$dX_8/dt = -0.545\ x_8^{0.596}x_{16}^{0.921}x_{25}^{0.078} + 43.113\ x_7^{0.949}x_9^{0.001}x_{15}^{0.972}x_{26}^{0.027}$$

$$dX_9/dt = -1.928\ x_9^{0.050}x_{17}^{0.972}x_{26}^{0.027} + 0.544\ x_8^{0.556}x_{10}^{0.041}x_{16}^{0.944}x_{27}^{0.055}$$

$$dX_{10}/dt = -0.337\ x_4^{0.034}x_{10}^{0.628}x_{18}^{0.892}x_{21}^{0.052}x_{27}^{0.055} + 1.795\ x_7^{0.0003}x_9^{0.048}x_{11}^{0.021}x_{17}^{0.973}x_{20}^{0.0003}x_{28}^{0.025}$$

$$dX_{11}/dt = -0.230\ x_{11}^{0.752}x_{19}^{0.971}x_{28}^{0.028} + 0.323\ x_2^{0.007}x_3^{0.006}x_{10}^{0.611}x_{18}^{0.991}x_{29}^{0.008}$$

Because there is no lumping of enzymes in this case, the dependencies of the rate of change of the metabolite pools upon the concentrations of the enzymes are all approximately first order i.e. to the power of 1.

Table 2.4: Eigenvalues of System Jacobian

-9292.25, -3070.91, -7.931, -0.070, -18.736, -19.872

Solving the linear system of Eqn.(33) gives a solution of the form in Eqn.(37) so that we obtain the two matrices containing the system sensitivities **L** and **M**. As expected, most of these sensitivities are relatively small indicating that many of the system components have little effect on each other or on the fluxes. A list of the largest positive and negative sensitivities is given in Table 2.6. The system flux and concentration control coefficients are illustrated more clearly in Figures 2.9 and 2.10 respectively.

Table 2.5: List of highest and lowest system sensitivities:

[i,j]	$\ln V_i / \ln X_j$	[i, j]	$\ln X_i / \ln \alpha_j$	[i,j]	$\ln X_i / \ln X_j$
[5,13]	-5.806	[8,9]	-24.14	[9,17]	-19.96
[7,8]	-1.353	[9,9]	-23.53	[9,18]	-2.59
[8,19]	-1.267	[7,9]	-23.48	[10,18]	-1.622
[7,13]	1.891	[6,9]	-20.36	[5,1]	6.537
[5,1]	5.379			[5,12]	9.151
[5,12]	6.531			[9,13]	15.26

By varying the input metabolite concentration, profiles of the intermediate metabolite concentrations and fluxes can be obtained and these are shown in Figures 2.11 and 2.12. The power-law approximation is exact at the operating point, but will be a poor approximation at regions far from the operating point. This is illustrated in Figure 2.13 which shows how the power-law approximation deviates from the behaviour resulting

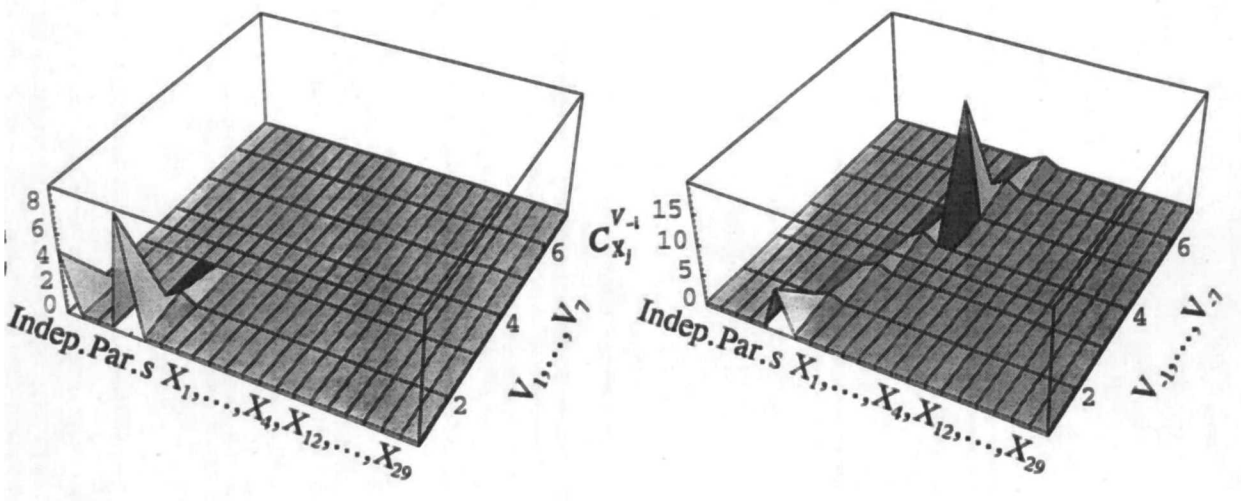


Figure 2.9: Flux control coefficients for full system.

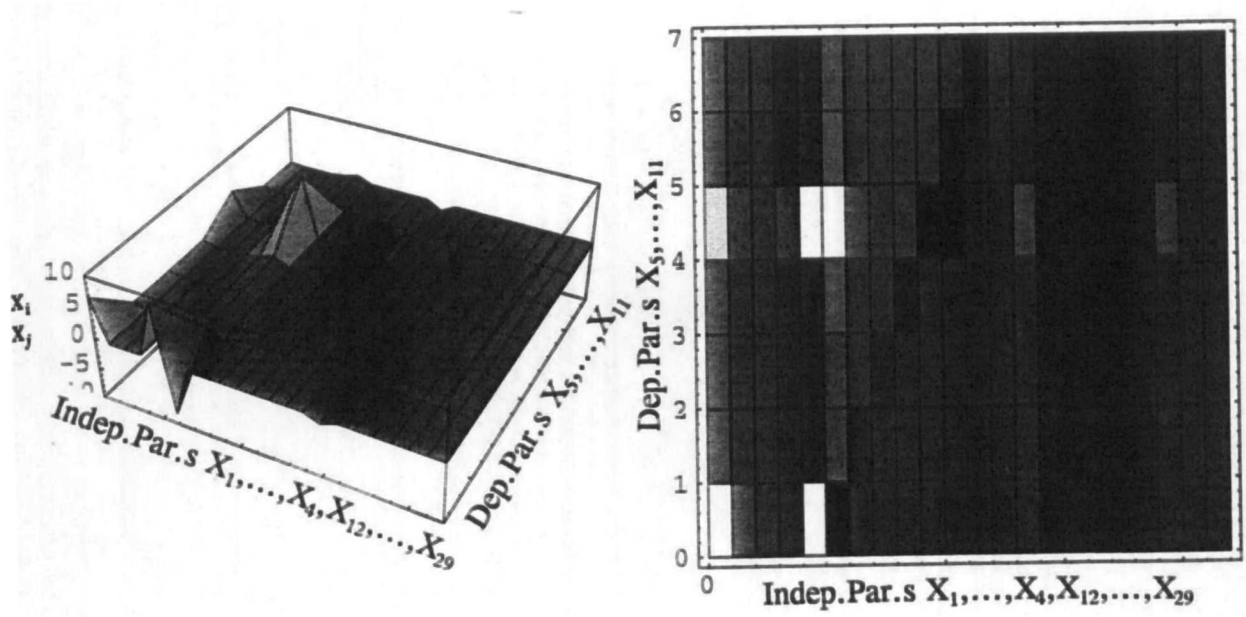


Figure 2.10: Concentration control coefficients for the full system. These show that the highest are C^{X5}_{X1}, C^{X5}_{E1}, C^{X9}_{E1} and C^{X9}_{E2}. This indicates that it is the first two enzymes in the pathway which exert the greatest control and the metabolites which are most affected are X_5 and X_9 (cis-diol and oxalocrotonate).

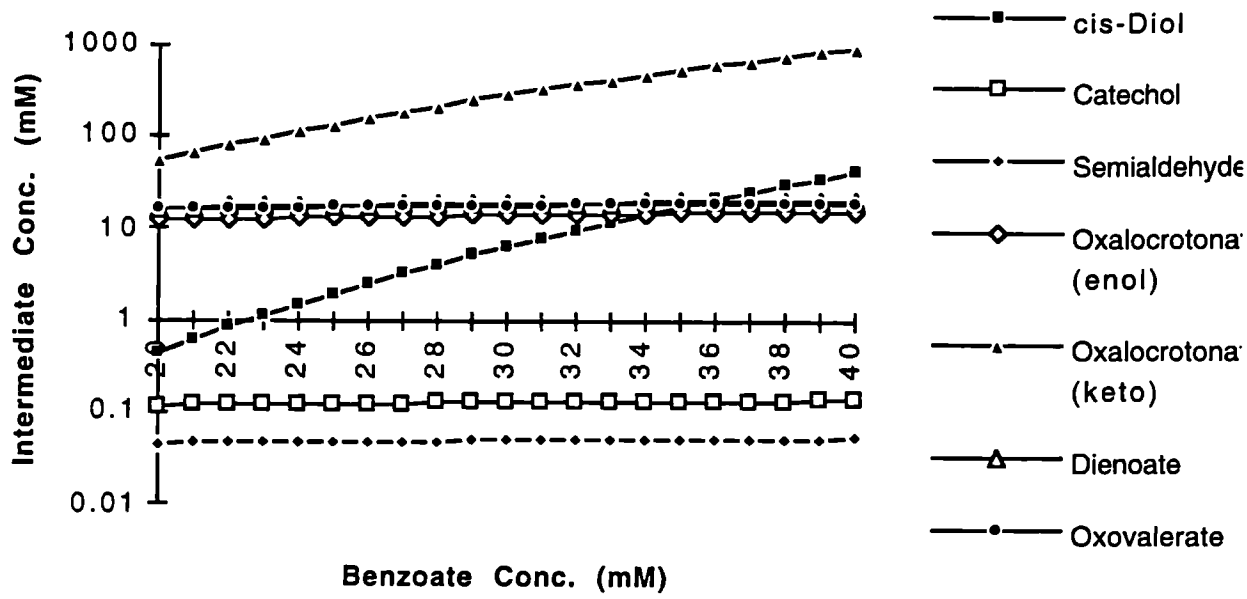


Figure 2.11: Predicted concentration profiles with changing input concentration.

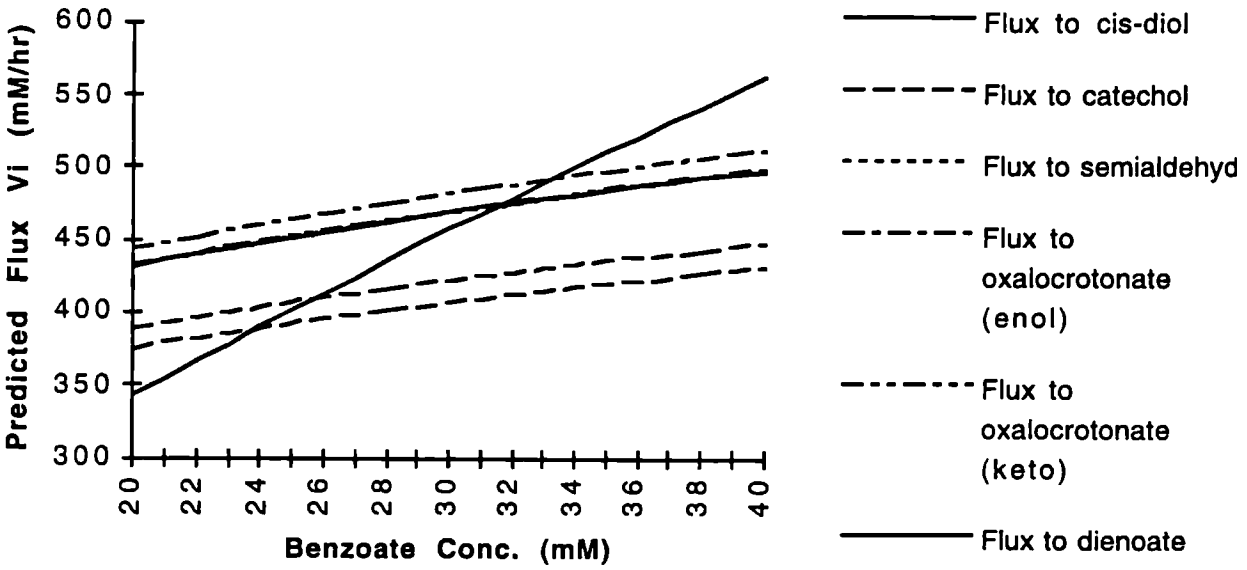


Figure 2.12: Predicted flux profiles with changing input concentration.

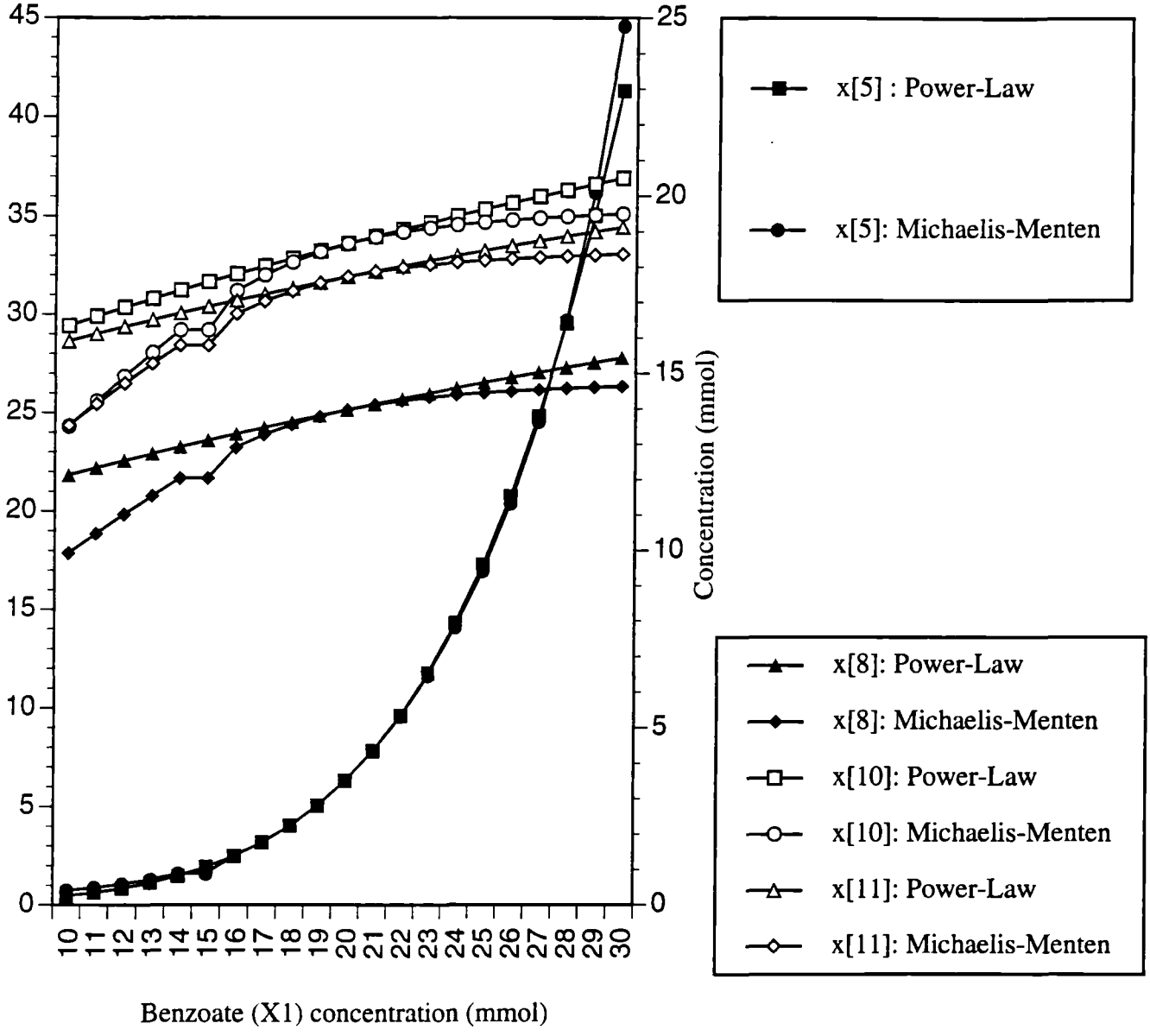


Figure 2.13 This figure shows the deviation of the power-law approximation from the Michaelis-Menten values in regions away from the operating point.

from Michaelis-Menten kinetics. It can be seen here that the predicted values of the parameters vary in their deviations from the Michaelis-Menten values. It can be seen that the deviation at either end of the range for cis-diol (X_5) is quite small whereas those for oxalocrotonate (X_8), (X_{10}) and (X_{11}) reach values of up to 15% at the lower end of the range.

Examining the sensitivity of the power-law system, the condition number (calculated using the F-norm) is 124.5. This is a measure of the relative error in x as a result of the relative error in A and b.

The results of the optimization of the system fluxes and concentrations are shown in Figure 2.14 and Figure 2.15. These are again displayed as 3-D surface plots and as density plots in order to show patterns of enzyme and intermediate concentrations.

2.7 Discussion

There are three issues to be examined, following from the results above: the behaviour of the two-step system, the behaviour of the full system and how the two compare.

The flux control coefficients of the two-step system indicate that the initial benzoate concentration and the activity of the first enzyme, the dioxygenase, have the strongest positive influence on the flux V_4, the flux which goes to increase the pool of catechol. The flux which goes to decrease this catechol pool, V_{-4}, is affected by the backwards first reaction and by the forwards second reaction. Not surprisingly, the same patterns are shown for the concentration control coefficients for catechol, although the coefficients differ quantitively.

Looking at the system parameters obtained when optimizing the concentrations of the intermediates and enzymes in the pathway, it can be seen that the system is quite insensitive to some parameters, for example, X_2, X_3 and E_4 (pyruvate, acetaldehyde and the lumped set of backward reactions). To maximize benzoate it is necessary that the activity of the dioxygenase be kept low. To maximize the catechol concentration, the dioxygenase activity should be as high as possible. The results for flux optimization are

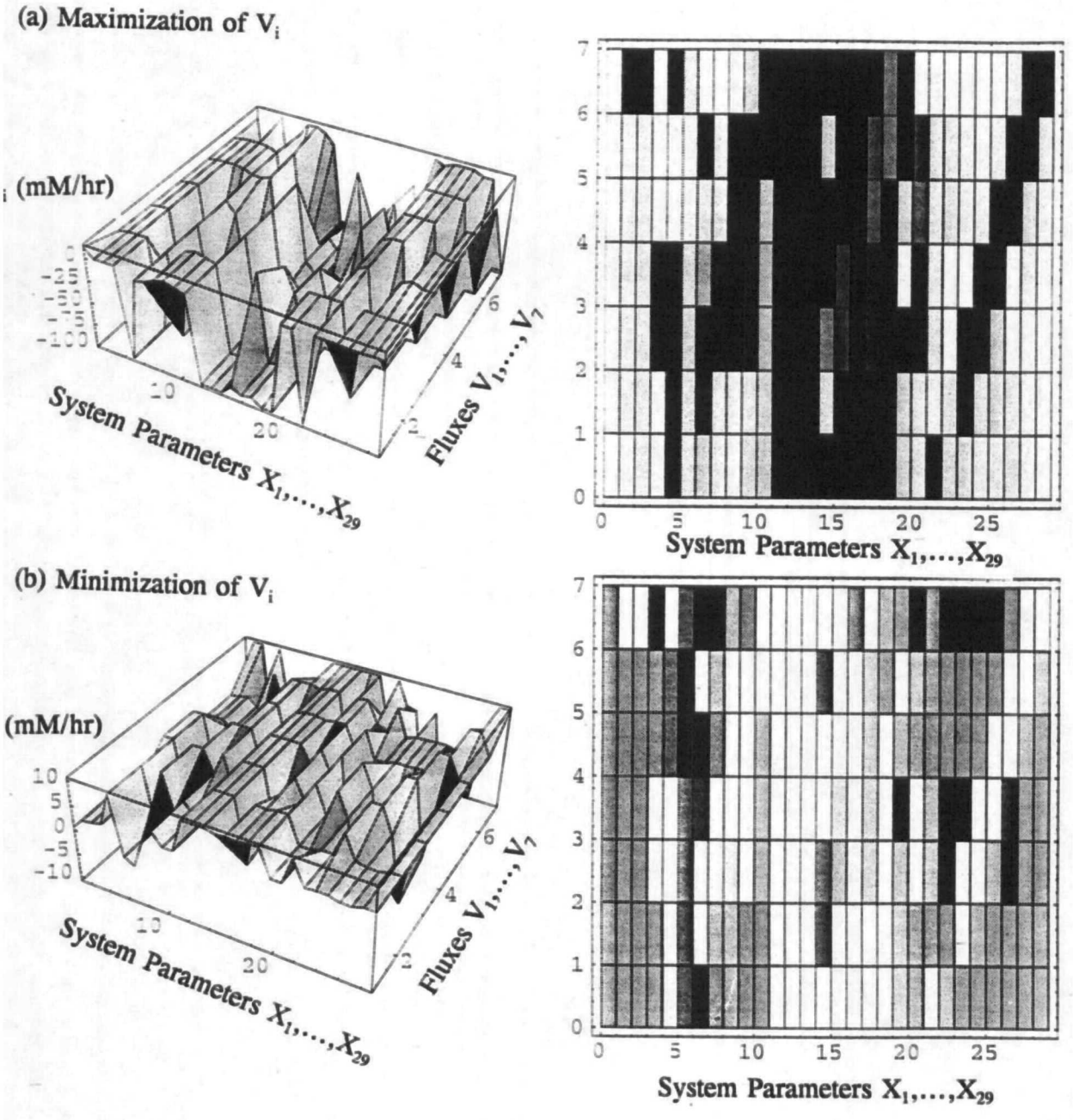


Figure 2.14: System parameters required for (a) the minimization of V_i and (b) the maximization of V_i for the full system.

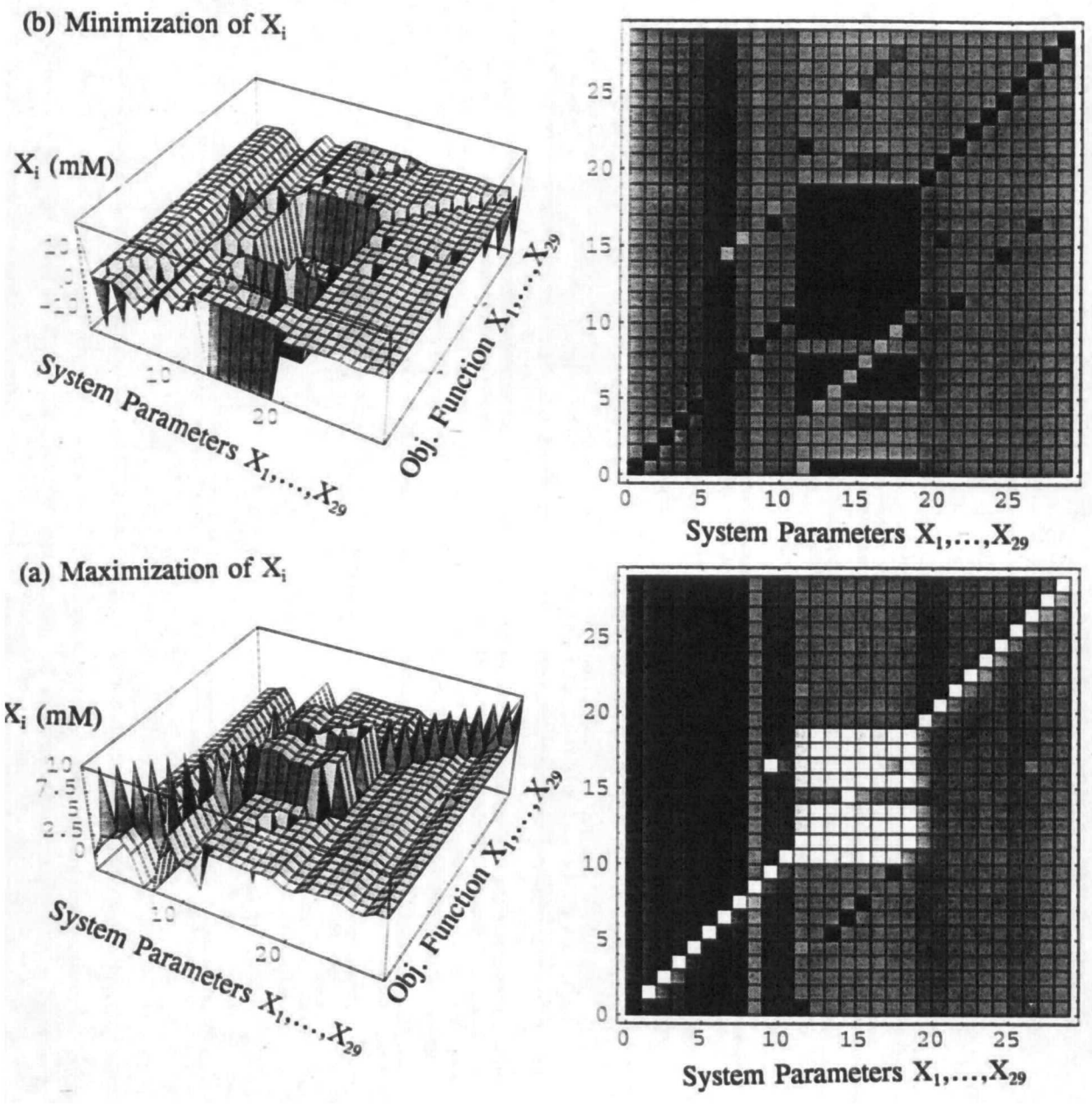


Figure 2.15: System parameters required for (a) the minimization of X_i and (b) the maximization of X_i for the full system.

slightly trivial: to maximize V_4, the initial benzoate concentration should be maximized, to minimize V_4, all of the system concentrations should be minimized. This is an indication that to obtain meaningful flux optimization results, it is necessary to further define the system.

The results obtained for the full TOL pathway are as expected for such a system. The concentrations of most of the metabolites in the pathway vary little on changing the input concentration. Only X_5 and X_9 show a significant change, increasing steadily with increasing benzoate concentration. While this would have been predicted for X_5 (being the second metabolite in the pathway and therefore directly affected by input levels) the profile for X_9 is more interesting as it is a result of the branched loop in the pathway.

Most of the rates increase gradually with increasing metabolite concentration, the exception being the flux to X_5 which rises significantly. Looking at the profiles of the concentrations and rates it is already possible to identify the bottlenecks in the pathway in qualitative terms. The increase in both the concentration of X_5 and the flux through it, followed by the lack of change in the variables immediately after it indicate that the first and second reactions in the pathway play a large part in controlling the rate through the pathway. The increase in the concentration of X_9 without a significant increase in its rate of formation indicates that this is another bottleneck.

While these profiles can indicate where in the pathway such bottlenecks occur, they do not quantify the extent of the control exerted or identify the rate and concentration controlling species. For this information we must examine the system sensitivities.

Table 5 lists the highest and lowest of the sensitivities of the dependent metabolites with respect to the independent metabolites, the sensitivities of the fluxes to the metabolite concentrations and the sensitivities of the metabolite concentrations to the rate constants. From the flux sensitivities we see that the first flux in the pathway is under the most control, in that it is the most sensitive to changes in the input concentration, and in the activities of enzymes catalyzing the first and second reactions in the pathway. The first enzyme in the pathway is the most rate-controlling although all of its control is exerted on the first flux in the pathway. The second enzyme in the pathway has a significant controlling effect on several of the pathway fluxes and may be more important in that respect.

Looking at the concentration control coefficients, we see that X_9 is under the most control being strongly influenced by changes in the levels of X_1 and enzymes X_{13}, X_{17}, X_{18} and to a lesser extent X_{21}. This is due to its position just before the branch point in the system.

Again the levels of the input concentration and the first and second enzymes in the pathway are seen to have a large effect on the concentration of X_5.

The sensitivity information on the rate constants points to the rate constant for the X_9 step in the pathway as having the greatest negative effect on several of the pathway intermediates.

One of the more notable points about the system sensitivities is that the concentration sensitivities are quite high. This is of interest in a system where many of the metabolites are toxic to the cell as it indicates that changes made to the system in an effort to improve the throughput flux may instead cause an increase in the concentrations of toxic intermediates, impairing or killing the cell.

The initial benzoate concentration has a strong positive controlling influence on all of the dependent system parameters as has the activity of the dioxygenase, E_1. This is evident from their control coefficients and is in agreement with the results obtained from the lumped two-step system. E_2 has a strong positive influence on all of the intermediate concentrations except that of benzoate. It can be seen from the density plot in Figure 2.11 that the enzyme following each intermediate has a negative control coefficient with respect to the concentration of that intermediate. The intermediate X_9, oxalocrotonate, displays a more complex pattern of control than the others and is affected both positively and negatively by several of the system components.

The flux control coefficients are more straightforward. All fluxes are negatively affected by the following enzyme, but to varying degrees: the flux to X_9 is the most strongly affected by its degrading enzyme, followed by the first flux in the pathway, and all the others are of a similar order of magnitude. The concentrations of benzoate and the cis-diol have the greatest effect on V_1.

Examining the optimization patterns of the metabolites in the pathway, the following trends appear. To minimize the concentration of each component in the pathway it is necessary to reduce the levels of X_5 and X_6 (the cis-diol and the semialdehyde) as much as possible. For minimization of X_6, X_7, and X_8, the enzymes E_1 to E_8 should be

minimized except for the enzyme which degrades the compound in question.

To maximize the system fluxes, E_1, E_2, and E_3 should in general be most active, X_5 should be high and X_6 low. The reverse holds for minimizing the fluxes. It should be noted that the conditions for minimizing and maximizing V_i are identical to those for maximizing and minimizing V_{-i} respectively as is physically reasonable.

As an example, two objective functions have been chosen for optimization. The first is the maximization of V_{-1}, the rate of degradation of benzoate; the second is the minimization of X_{11}, the toxic by- product, formate. The results for the maximization of benzoate degradation indicate that the strategy should be to: (a) increase the benzoate concentration in the cell by increasing the rate of transport into the cell and by increasing the cell's tolerance to high benzoate concentrations; (b) minimize the concentrations of the end-products, and (c) over-express the levels of the enzymes catalyzing the forward reactions in the pathway. The minimization of the by-product formate (X_4) at constant benzoate concentration indicates that the end- product concentrations should be minimized. Also the level of enzyme X_{20} should be decreased while that of X_{29} should be increased. This result indicates that increasing the level of all of the enzymes in the pathway simultaneously will not necessarily result in increased flux or improved performance. The positive rate-controlling enzymes must be identified as above, and then cloned and over-expressed individually.

This program proposes a structured way of examining the behaviour of metabolic pathways. Data can be either from in vitro kinetic data or in vivo fermentation data so that it is of use in preliminary studies and fermentation design as well as fermentation studies and process optimization. It has the advantage of examining processes at a basic metabolic level so that it provides direction for genetic engineering programs.

The sensitivity analysis pinpoints the rate-controlling steps and shows where efforts should be focused on removing bottlenecks so that yields can be improved. The optimization routine indicates the set of operating parameters required for optimum performance of the system as well as giving upper and lower bounds for productivity and metabolite levels.

3. MATERIALS AND METHODS

3.0 Introduction

The models in the previous chapter demand that a broad range of analytical techniques be employed to yield the appropriate data. The primary requirement is that the pathway intermediate concentrations be measured over a period of time to determine the transient concentrations. This lead to several issues: firstly the fermentation must provide sufficient density of cells to carry the plasmid encoding the pathway and these cells must be of an appropriate metabolic state to express the enzymes of the pathway and to supply the necessary cofactors for the pathway reactions. The levels of these enzymes is dependent upon the copy number of the plasmid as well as the strength of the pathway promoter. The rate at which the degradation of the pathway substrate proceeds will be dependent on the substrate concentration as well as the enzyme levels so it is desirable to have initial levels as high as possible without causing damage to the cell.
The analyses which must be performed on the system can be divided into two types: those which will measure global fermentation parameters which are bulk properties of the system and those which will measure pathway parameters which are specific to the pathway of interest. Into this first group fall such parameters as pH, temperature, dissolved oxygen concentration, CER, OUR, biomass, total carbon and total nitrogen whereas the second group refers to specific metabolite concentrations and enzyme activities.
Initial shake flask experiments were used for media optimization. The final fermentation protocols was chosen on the basis of initial experiments as described in Section 4.1.

3.1 Strains

The strain used is *Escherichia coli* JM107 which has been transformed with a number of different plasmids. The first of these is pQR150. This plasmid is a derivative of pBGS18 into which has been inserted the meta-cleavage pathway of the TOL plasmid

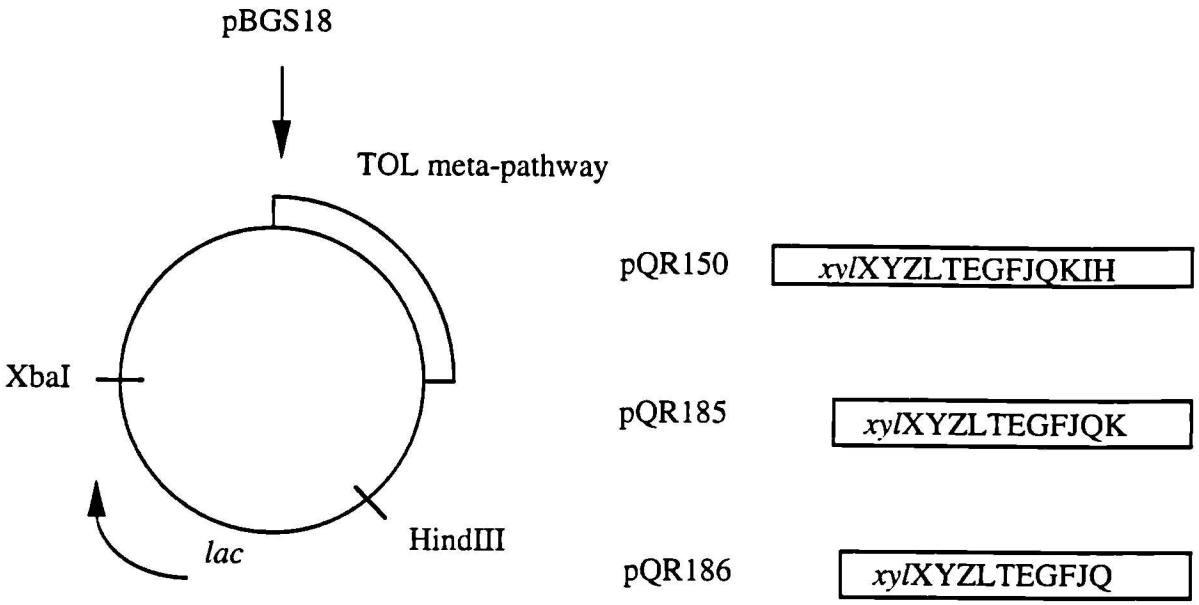
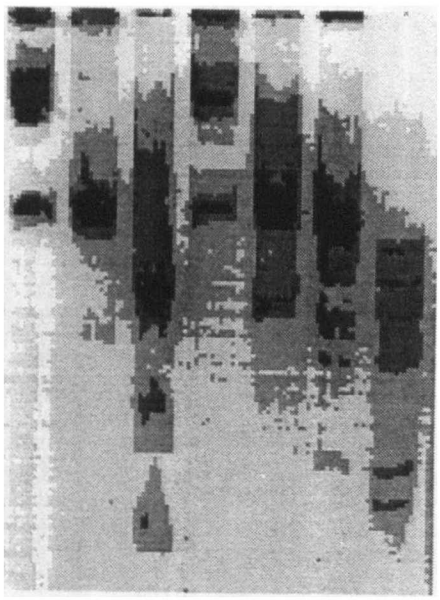


Figure 3.1: The construction of the plasmid pQR150 and the gene order of the plasmids pQR185 and pQR186.



1. 1+ Control.
2. 1 + Xba + HindIII
3. 1+ XhoI
4. 2 + Control
5. 2 + Xba + HindIII
6. 2+ XhoI
7. λ HindIII marker

Figure 3.2: Miniprep of pQR150 showing presence of plasmid. Number 2 shows two bands for the Xba HindIII digest in lane 5 as expected.

from *Pseudomonas putida* (see Figure 3.1) with the natural promoter replaced by the *tac* promoter. Presence of the plasmid was confirmed initially by DNA mini-prep and gel (see Figure 3.2), where the presence of two bands in lane 5 indicates the correct conformation, and subsequently by catechol dioxygenase activity, that is, colonies containing the plasmid will convert colourless catechol solutions into bright yellow coloured 2-hydroxymuconic semialdehyde.

The other plasmids studied, pQR185 and pQR186 were truncations of this pathway. The gene orders for these plasmids are also shown in Figure 3.1.

Cultures were plated out on M9 minimal media agar plates (see below) and incubated at 37°C until growth was apparent (approximately 12-18 hours). Plates were then stored at 4°C and subcultured every 10 days. Stock cultures were stored in a 1:1 glycerol:media mixture at -80°C.

3.2 Media

Two types of media were used: one complex (LB) and one defined (M9). LB media was composed of: tryptone (20 g/l), yeast extract (10 g/l), NaCl (10 g/l), with glucose (30 g/l) as carbon source.

M9 media of the following formulation was used: Na_2HPO_4 (6 g/l), KH_2PO_4 (3 g/l), NaCl (0.5 g/l), NH_4Cl (6 g/l). This was stored as a 10 times concentration stock solution for shake flask cultures and freshly made up for 1.5 l fermentations. To this was added $CaCl_2$ (10 mM) and $MgSO_4$ (100 mM), stored separately as a 100 times stock solution. Agar (1.5 %) was added for making up slopes and plates.

This media was supplemented either with thiamine or with 0.1% yeast extract. Positive pressure for retention of plasmid-bearing cells was exerted by the addition of 25 mg/l kanamycin and the meta-cleavage pathway was induced by the addition of IPTG. Both of these were sterile filtered through a 0.4 μm filter (Gelman Sciences, Ann Arbor, MI, USA) and added after sterilisation.

The carbon sources used were glucose (10 g/l), glycerol (10 g/l) and benzoic acid (2-20 mM). 1 Mm IPTG was used for induction of the *tac* operon of the meta-cleavage pathway.

Yeast extract was obtained from Difco, E.Moseley, Surrey. All other chemicals were obtained from Sigma Chemicals Inc, Poole, Dorset, U.K. A range of substrates was supplied to the pathway. The primary substrates were benzoic acid and sodium benzoate. Also used were p-toluic acid (4-methyl benzoate) and m-toluic acid (3-methyl benzoate). When difficulties in dissolving the substrates occurred, 1% ethanol was used to solubilize them before adding to the medium.

3.3 Fermentation

3.3.1 Shake Flask Fermentations

Agar plates were prepared from stock cultures and incubated for one day at 37°C after which colonies were tested for presence of the plasmid with catechol. These plates were used to prepare overnight cultures in M9 medium, 1 ml of which was used to inoculate the 100ml shake flask cultures. These were incubated for 12 hours at 37°C and 200 rpm in an orbital shaker.

3.3.2 Batch Fermentations

Batch fermentations were carried out in a 2 L LH fermenter (LH Engineering, Reading, Berkshire). The fermenter configuration process overview is outlined in Figure 3.3. A combination pH electrode and a polarographic oxygen electrode, both manufactured by Ingold, were used to measure pH and dissolved oxygen tension, respectively. Temperature was also measured and off-gas analysis of CO_2, O_2 and N_2 was carried out by a VG MM 8-80 mass spectrometer. Temperature was controlled at 37°C and pH was controlled at 7.00 by the addition of 2 M NaOH using TCS controllers (Worthing, UK). DOT was kept above 40% by controlling the agitation rate. Data was acquired and logged by either of two data acquisition systems: the first was a DEC PDP 11-73 operating the Bio-i (Biotechnology Computer Systems) and the second was RT-DAS (Real Time Data Acquisition Systems). Two methods were used to control foaming: one

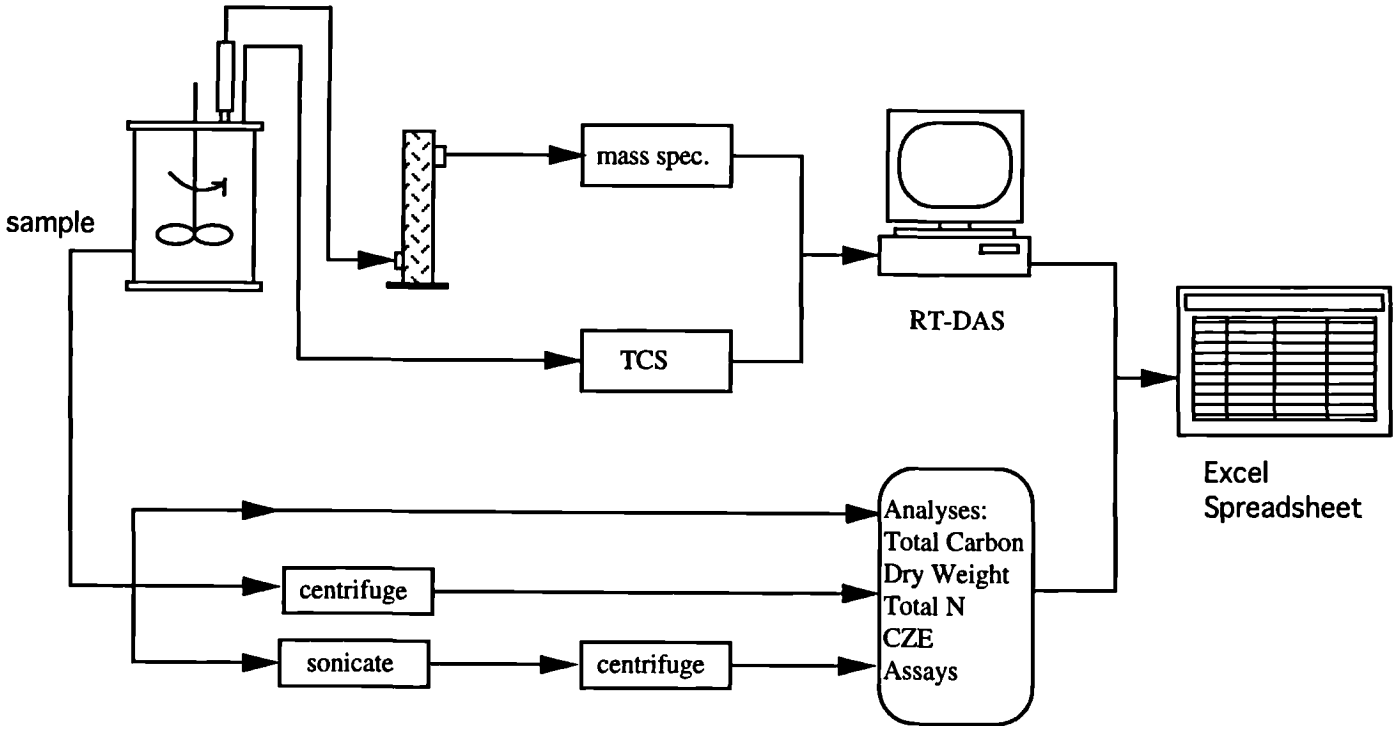


Figure 3.3: Fermenter configuration and process overview

was the addition of 0.5 ml/l PPG, the other, used to avoid the interference of PPG with analysis was the use of a glass wool trap between the fermenter and the outlet filter as shown in Figure 3.3.

Before sterilisation, the pH probe was calibrated to pHs 7 and 4 using standard buffers. The DOT probe was calibrated to 100% air saturation by bubbling air through the medium, and to 0% by bubbling nitrogen through and waiting for the reading to stabilise. Having calibrated the probes, all inlet and outlet lines except the gas outlet were clamped shut and all sterilisable connecter ends and electrical connections covered with cotton wool and aluminium foil. The fermenter was then filled with 1.5 l of medium and the clamps closed loosely. The assembled fermenter was then sterilised by autoclaving at 141°C for 20 minutes. The base addition cylinder was similarly sterilised. On cooling the necessary attachments were made to the stirrer, pH, DOT and temperature controllers and the heating and cooling fingers. The NaOH addition cylinder was filled with 2 M NaOH and the stericonnectors joined aseptically. The controllers were then switched on to allow the fermenter to reach a temperature of 37°C and a pH of 7. The DOT probe was then re-calibrated to 0 % and 100 % by bubbling nitrogen and air through, respectively. The air inlet was attached to a compressed air supply and the air outlet to a mass spectrometer line.

A 1 ml overnight culture was used to inoculate a 150 ml shake flask culture in an inoculation flask. This was added to the fermenter at mid-exponential phase of growth. Just after inoculation, kanamycin was sterile filtered and added to a concentration of 25 mg/l. The initial stirrer speed was 400 rpm and the initial air flow rate 0.5 vvm. DOT was kept above 40 % by manually controlling first the stirrer speed (up to a maximum of 1000 rpm) and then the air flow rate.

Samples were taken hourly and the temperature immediately reduced to 4°C by immersion in ice to minimize metabolic activity during sample processing.

Several fermentation and feeding regimes were used, as described below:

(i) Batch fermentation: M9 + 10 g/l glycerol + 10 mM benzoate
-pathway not induced

(ii) as in (i) but induced at t=0 with IPTG

(iii) Fed-batch: M9 + 10 g/l glycerol
-feed of benzoate + glycerol at stationary phase, pathway not induced

(iv) as in (iii) but induced with IPTG upon addition of benzoate.

3.4 Analyses

3.4.1 Spectrophotometric Measurements

Spectrophotometric measurements were carried out on the whole broth at 560 nm to measure the optical density (O.D.) and on the supernatant at 320 nm to measure the concentration of hydroxymuconic semi-aldehyde. Samples were also scanned from 210 to 700 nm. The spectrophotometer used was a Beckman DU-700. A deionised water blank was used to zero the readings. Samples were diluted to fall within the linear range of the instrument.

3.4.2 Measurement of Biomass

1.5 ml eppendorfs were dried for at least 24 hr and weighed before analysis. 3 x 1 ml of sample was placed in these dried, pre-weighed eppendorfs and spun down for 2 minutes at 10,000 rpm in a Beckman microcentrifuge. The supernatant was taken off and retained, a second aliquot was added to each eppendorf and spun down again. Supernatant was again retained for analysis and the biomass was washed with 1ml deionized water, resuspended and spun down. This washing process was repeated. The eppendorfs with biomass were dried in a 100°C oven for 24 hours and then weighed. The supernatant was stored at -20°C before analysis.

3.4.3 Presence of Plasmid

Presence of the plasmid and activity of the pathway enzymes was monitored by placing catechol into a small amount of the fermentation broth. A bright yellow colour indicated the presence of an active catechol-2,3-dioxygenase.

3.4.4 Assays

Formate, pyruvate and glycerol were assayed using commercially available kits (Boehringer-Mannheim, Lewes, E.Sussex).

3.4.5 Total Organic Carbon

Total organic carbon content of the broth and supernatant was measured by a Shimadzu TOC-5050. This instrument works by introducing an aliquot of sample into a total carbon combustion tube which is filled with oxidation catalyst and heated to 680°C. Carrier gas is supplied into the combustion tube at a controlled rate and this carriers the CO_2 resulting from the combustion of the total carbon in the sample into an inorganic carbon reaction vessel and then it is cooled and dried by a humidifier. It is then sent through a halogen scrubber into a sample cell set in a non-dispersive infrared gas analyzer (NDIR) where CO_2 is detected. The NDIR outputs a detection signal which generates a peak whose area is calculated by a data processor.
Both broth and supernatant were diluted 1:10 to fall within the optimum range of the instrument. Calibration curves were obtained by using potassium hydrogen phthalate in concentrations of 0.1 0.5, and 1 C g/l as standard for total carbon and sodium hydrogen carbonate and sodium carbonate in concentrations of 0.1, 0.5 and 1 C g/l as standard for inorganic carbon. The sample was sparged with nitrogen for 1 minute before 4 washes of the system were carried out. Three repeat injections were routinely made with a maximum of 5 if a coefficient of variance of 2 % was not initially achieved.

3.4.6 Total Nitrogen

Total nitrogen was measured using a LECO nitrogen analyzer. This instrument analyses nitrogen in three phases: purge, burn and analyse. In the purge phase, the encapsulated sample is placed in the loading head and purged of any atmospheric gases that have entered during sample loading. The ballast volume and gas lines are also purged.

During burn phase the sample is dropped into an 850°C furnace and flushed with pure oxygen. The products of combustion, CO_2, H_2O, N_2 and NO_x are passed through a thermoelectric cooler to remove most of the water and collected in the ballast volume. In the analyse phase, a piston on the ballast is forced down and 10cc of the sample is collected through the loop of the doser valve. The sample is swept through hot copper to remove oxygen and to convert NO_x to N_2, then through Lecosorb and Anhydrone to remove CO_2 and H_2O respectively. The N_2 remaining is measured by a thermal conductivity cell which is based on a Wheatstone bridge which is balanced with helium. Readings are given to 0.001 % with an accuracy of +/- 0.003 %, when the instrument is operating in its optimal range of detection. This is in the range 0.01 g- 1.0 g nitrogen. For amounts below this, the accuracy will decrease (to +/- 2 % for amounts of 1 mg or below). This means that there is an optimum accuracy which can be obtained with dilute samples by concentrating (e.g. by evaporation).
At least three air blanks measurements were made before each use, until the readings were within +/- 0.003 % of each other. EDTA was used as a standard.

3.4.7 Capillary Zone Electrophoresis

3.4.7.0 Introduction

Capillary zone electrophoresis (CZE) is an analytical technique which separates components in a mixture using electrophoresis, then measures the absorbance of the components by UV spectrophotometry. The result is a quantitative output similar to a chromatogram called an electropherogram, but obtained by means of a different separation principle.

3.4.7.1 Principle of Operation

The configuration of a typical CZE system is shown in Figure 3.4. It consists of a silica-coated capillary, whose inner diameter can range from 20-200 μm though it is

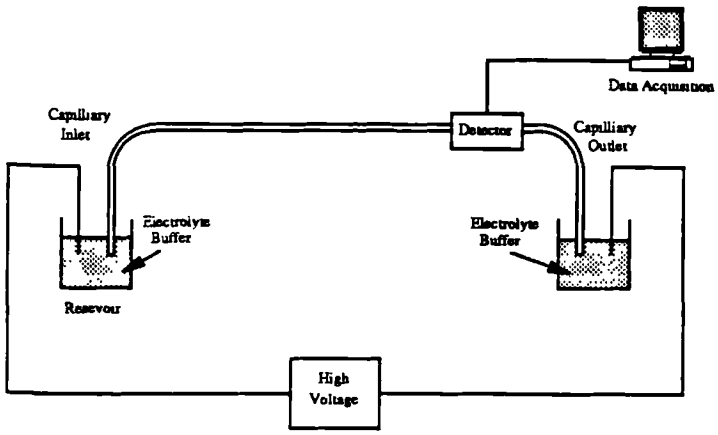


Figure 3.4: Configuration of capillary zone electrophoresis system

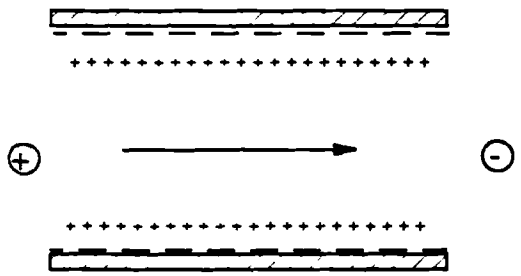


Figure 3.5: Charge effects at capillary surface

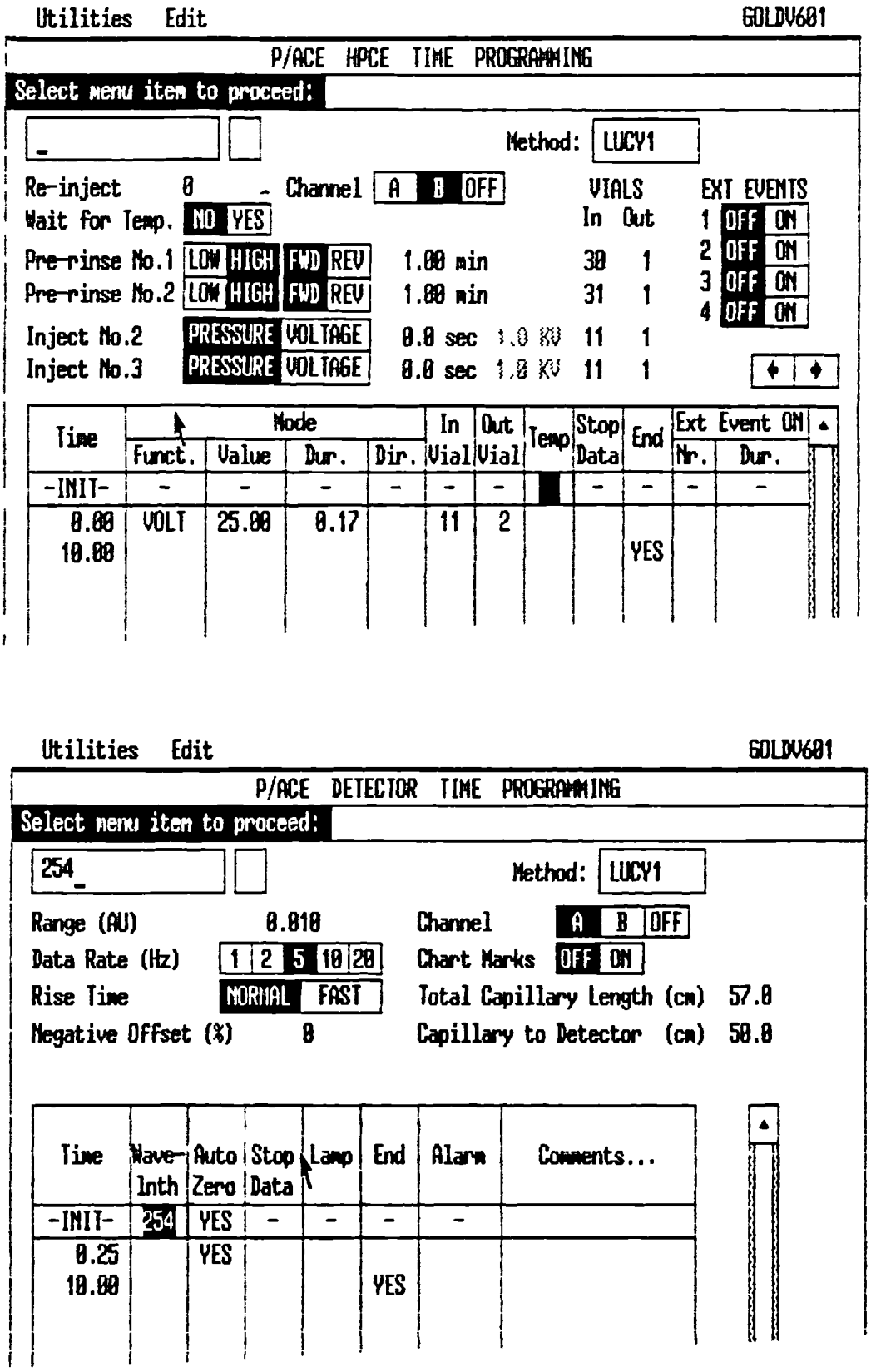


Figure 3.6 Method table for PACE and detector unit showing conditions for separation of intermediates.

typically 50 or 75 μm inner diameter and ranging from 7 to 100 cm in length, which is filled with buffer and each end of which is immersed in buffer. After injecting an amount of sample, a high voltage, up to 30 kV is applied across the capillary and this causes the electrophoretic movement of the buffer as a whole and of the sample past a detector window where a UV detector monitors the absorbance and the results are logged.

The process that drives the separation is electroosmosis. This occurs because of the surface charge properties of the ionizable silanol groups of the fused silica capillaries which are in contact with the buffer. When a voltage is applied across the capillary the wall becomes negatively charged to a degree which is controlled by the pH of the buffer. This negatively charged wall attracts positive ions from the buffer, creating an electrical double layer (see Figure 3.5), so that when a voltage is applied across the capillary, cations in the solution move towards the cathode, in a process known as electroosmotic flow. The electroosmotic flow is defined by the equation

$$v_{eo} = \frac{\epsilon \zeta}{4 \pi \eta} E$$

where ϵ is the dielectric constant, η is the viscosity and ζ is the zeta potential. The zeta potential is related to the inverse of the charge per unit surface area, the number of valence electrons and the square root of the electrolyte concentration. Since it is an inverse relationship, increasing the concentration of the electrolyte decreases the electroosmotic flow.

In CZE, when the voltage is applied across the capillary, the components in the sample begin to separate into bands according to their electrophoretic mobility μ_{ep} which is defined by

$$\mu_{ep} = \frac{q}{6 \pi \eta R}$$

where q is the net charge, R is the Stokes radius and η is the viscosity. That is, the rate at which a compound will move is dependent upon its charge:mass ratio where the net charge is usually dependent on the pH of the buffer.

3.4.7.2 CZE Parameters

Analysis was carried out on a Beckman P/ACE 2000, controlled by System Gold software. the capillaries used were 50 um i.d and 50 cm length to detector (57 cm overall). The analysis parameters were entered as a method and sample table on System Gold. The method and sample table format and the system parameters are shown in Figure 3.6. The buffers used were as follows: borate 50 mM pH 7.5, phthalate 10 mM, pH 8.2.

3.4.7.3 Media

There will be always be a background caused by the media components. This may vary over the course of a fermentation as media components are utilized by the microorganism. For this reason it is preferable to be able to identify this changing background rather than simply subtract the initial baseline from all subsequent electropherograms. An electropherogram of M9 media whose formulation is given in Section 3.2, identifying the contribution from the individual media components is shown in Figure 3.7.

3.4.7.4 Standards

The compounds which were available in pure form and which could be used as standards were: benzoate, catechol, pyruvate, formate, acetaldehyde, cis-diol and oxalocrotonate. The wavelength at which each compound showed maximum absorbance was determined by spectrophometric scanning and this determined what wavelength was used in the CZE UV detector. The separation of benzoate, catechol, cis-diol and oxalocrotonate in M9 media was obtained using a borate buffer system and is shown in Figure 3.8 where all compounds are at 10 mM concentration. Calibration curves were obtained by running standards at concentrations of 1 mM, 5 mM and 10 mM.
The separation of pyruvate, formate and acetaldehyde was carried out using a phthalate buffer. Calibration curves were obtained by running concentrations of 1, 5 and 10 mM.

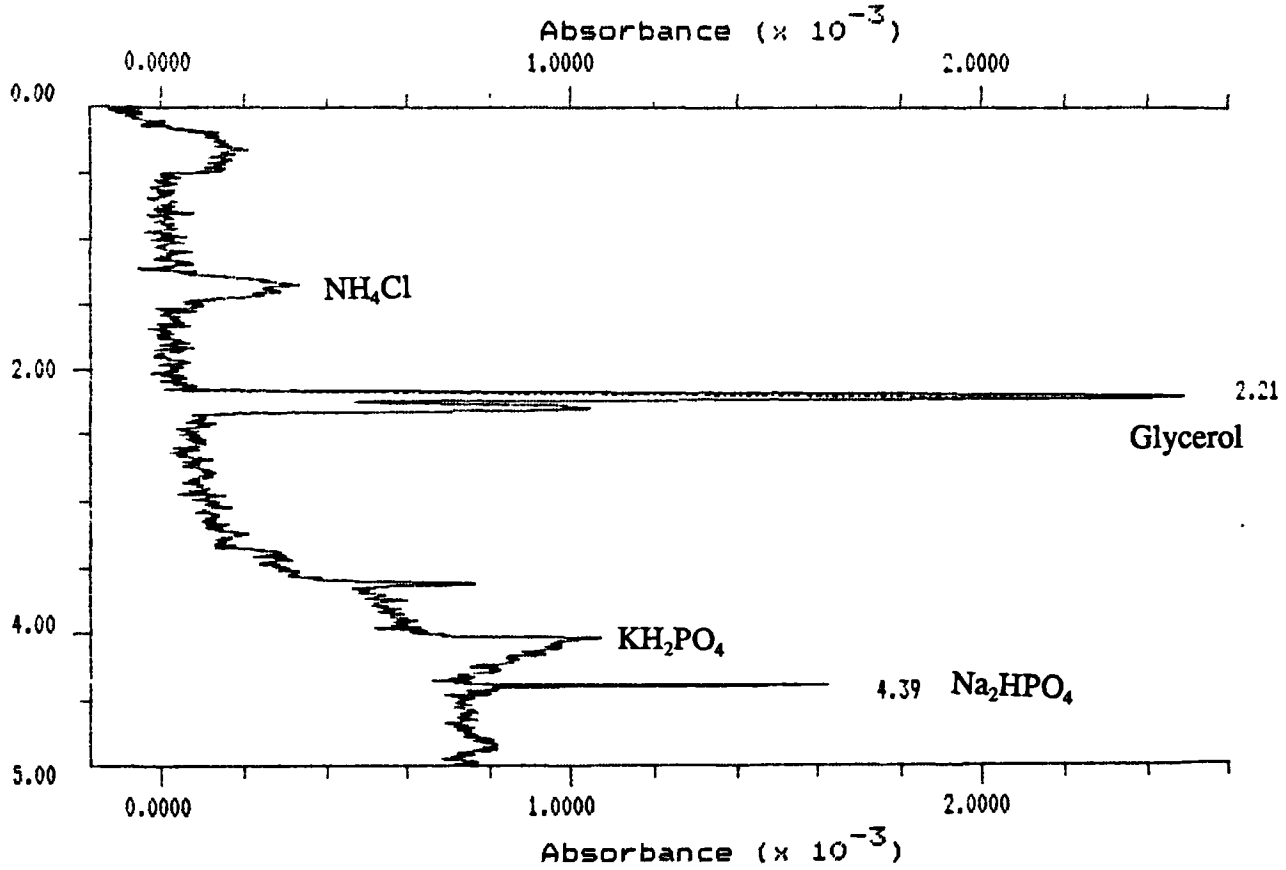


Figure 3.7: Electropherogram showing effect of presence of M9 media components on the baseline for the separation conditions shown in Figure 3.6 and using a 10 mM, pH 7.2 borate buffer system.

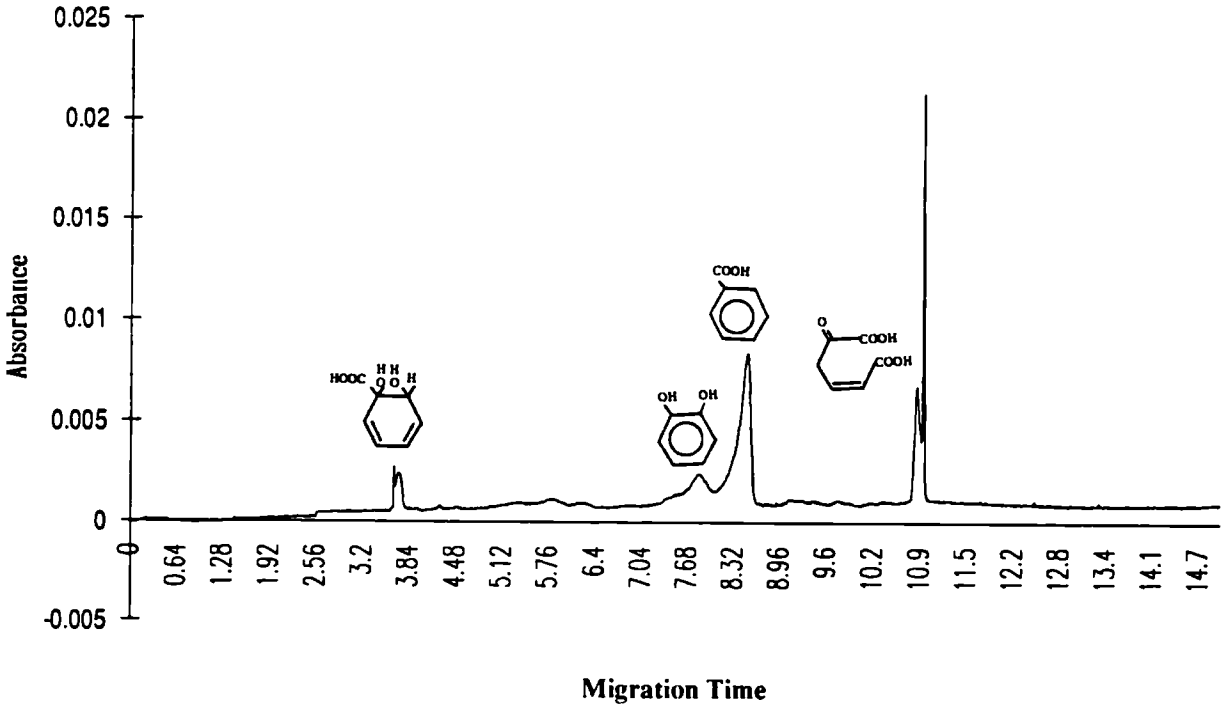
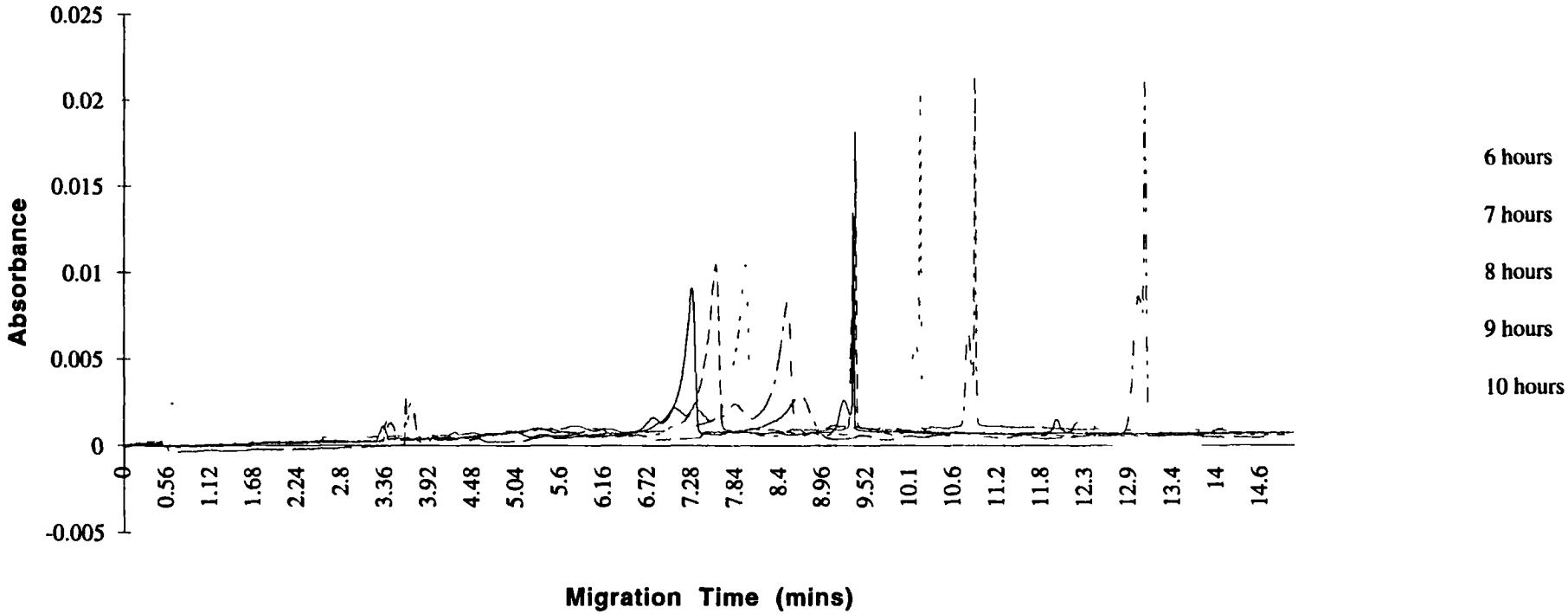


Figure 3.8: Electropherogram showing separation of TOL pathway intermediates using the separation conditions shown in Figure 3.6 and using a 10 mM, pH 7.2 borate buffer system.

Figure 3.9: Peak migration due to change in ionic composition with fermentation time can be seen in the movement of the reference peaks in this electropherogram.



3.4.7.5 Effect of Changing Composition

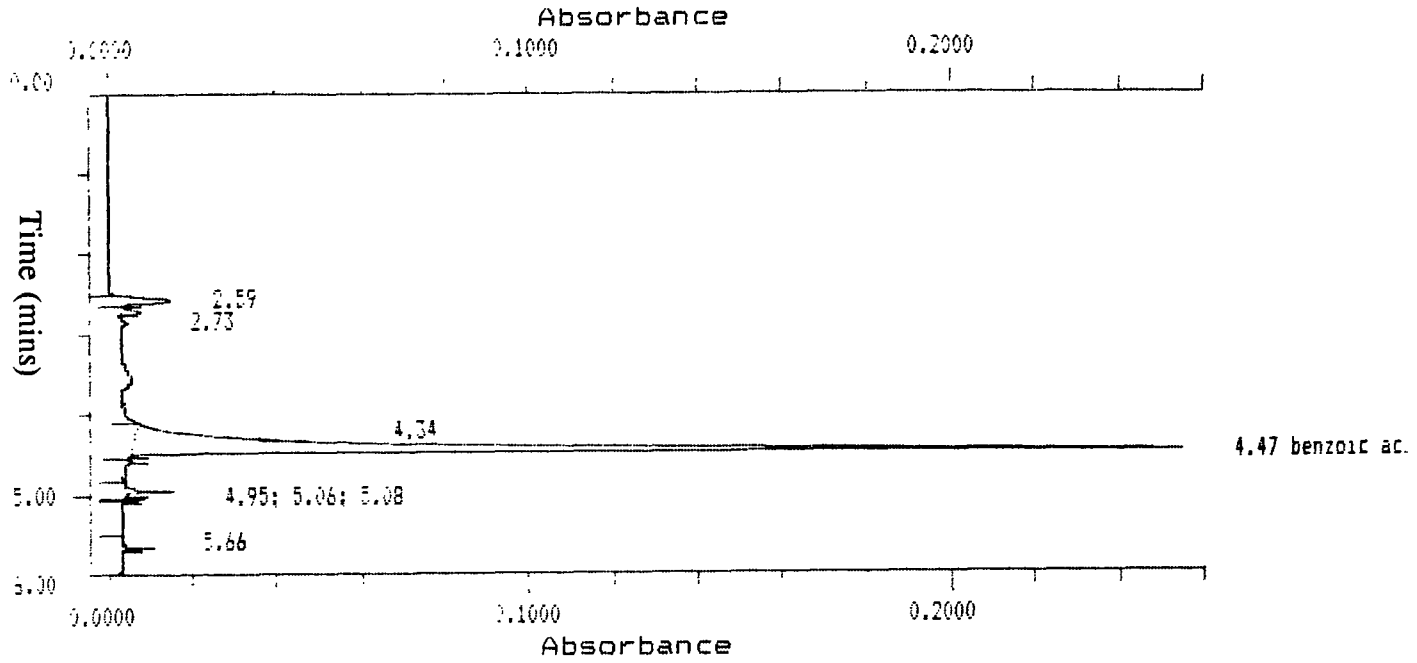
Over the course of a fermentation the ionic composition of the samples taken will change. This is due to the utilisation of media components by the cells, the excretion of metabolites such as organic acids by the cells, the presence of proteins and other polypeptides from cell lysis, and the degradation of benzoate and the production of pathway intermediates by the TOL meta-cleavage pathway. This change in ionic composition will cause a shift in the migration times of the sample components as seen in Figure 3.9. As the benzoate peak is easily identifiable throughout the fermentation, this peak is used as a marker and all other peaks are assigned with respect to their distance from the benzoate peak.

3.4.7.6 Effect of Antifoam

The antifoam agent used in these experiments was PPG. PPG acts as a surfactant and so breaks down the foam formed by the culture. The effect of addition of PPG to a fermentation on the CZE analysis is quite noticeable because of its surfactant nature. This is complicated by the fact that it is difficult to relate the amount of PPG remaining in a sample after the sample has been stored and processed, with that originally in the fermentation broth. Only if a colloidal emulsion can be maintained in the sample throughout processing and up to the time of analysis, can the PPG concentration be assumed to be equal.

Figure 3.10 shows the effect of adding 0.5 ml/l PPG to M9 medium, when the PPG is evenly suspended and the sample is analysed immediately, and when the sample has been allowed to stand for 1 hr and has been centrifuged at 10,000 rpm for 10 minutes.

M9 media with 10 mM benzoate



(b) M9 media with 10 mM benzoate + 0.5 ml/l PPG

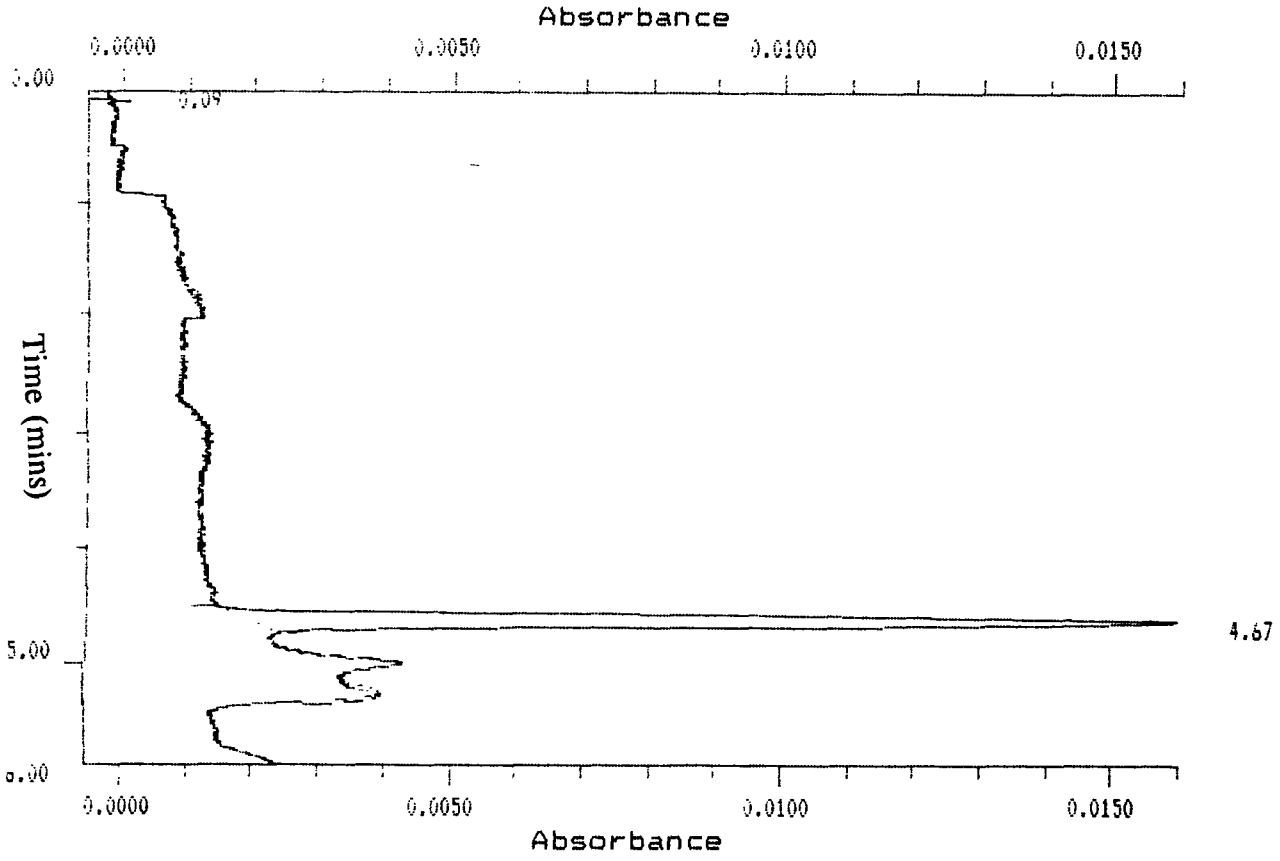


Figure 3.10: Electropherogram showing the effect of the antifoam PPG on the electropherogram of an M9/benzoate solution. The effects which can be seen are: poor baseline, shift in migration time and presence of unidentifiable peaks.

4. FERMENTATION

4.0 Introduction

The purpose of the fermentation in this case is to provide sufficient information to enable the calculation of the fluxes and of the control coefficients of the meta-cleavage pathway. As these control coefficients will be calculated on the basis of transient intermediate concentrations, it is necessary that the pathway produces and excretes to the medium intermediates in concentrations great enough for analysis. To do this, it is first necessary to define the system in terms of the pathways to be considered, the cellular environment of the pathway in terms of the host organism, and the process environment in which the pathway will operate. It is this last consideration which will be examined in this chapter which will define the fermentation conditions and determine the concentrations of the pathway intermediates.

The pathway engineering objective is to maximize the flux through the pathway i.e. maximize the rate of degradation of benzoate. Without attempting to alter the pathway by genetic means there are many process characteristics which can be optimized. These fall into the broad categories of media, substrate feed and fermentation protocol.

The first requirement of the medium is that it provides sufficient carbon and nitrogen for the required cell density. It must also provide salts and trace elements. Complex media such as nutrient broth and LB media will usually result in higher growth rates as amino acids and peptides are being supplied and the cell does not need to synthesize them. However the cell may also use the amino acids as a carbon source so that it is difficult to perform material balances. Minimal media overcomes this problem by supplying only defined components to the cell. This also has advantages in that it provides a low noise background for analysis. Choice of carbon source can greatly influence the growth characteristics of the cell. One of the most common carbon sources used is glucose. This is very easily and rapidly metabolised by *E.coli*, but causes the production of acetate which represses growth.

Under the heading of fermentation protocol comes the issues of whether to run in batch, fed-batch or continuous culture, fermentation volume, size of inoculum, point of

induction, point of substrate feed. These issues are primarily decided on the basis of the purpose of the fermentation and what is to be gained from it.

The stage in the fermentation at which the pathway substrate is introduced is also crucial: whether it is at the stage of inoculation, at the beginning, middle or end of logarithmic growth, or at the beginning of stationary phase will have different effects on the cell physiology. It must also be decided whether the substrate is to be pulse or continuously fed to the culture. A pulse feed will expose the cells to higher concentrations of the toxic substrate but may increase the rate of degradation.

All of the above considerations must be taken into account when deciding on fermentation conditions and were used to guide initial experiments.

4.1 Choice of Media

The choice of media was guided by the following considerations: it must give good, rapid growth of *E.coli* JM107 and it must allow for easy analysis. The M9 minimal media described in Section 3.2 had been used in the development of the strain and this was the basic media which was chosen for use as it had historically given good growth of JM107 and because it was a defined media which would simplify the analysis as well as carbon balancing.

One possible carbon source for use was benzoate as this is degraded to form pyruvate which is then passed to the central carbon metabolism. This would have had the effect of linking cell growth directly to the TOL pathway performance. To examine its performance as a carbon source a series of shake flask experiments were conducted, the results of which can be seen in Table 4.1. This shows that benzoate cannot act as a sole carbon source for the growth of *E.coli* JM107 pQR150.

This indicated that an additional carbon source was necessary for growth of the organism with benzoate supplied as a substrate for the pathway. Of the candidates for this carbon source (glucose, lactose, galactose, glycerol) glycerol was chosen.

Because benzoate is toxic to *E.coli* (benzoic acid is a commonly used preservative and anti-bacterial agent) it was also necessary to determine what is the allowable concentration for growth of *E.coli*. For subsequent analysis it was necessary to use the

highest possible concentration of benzoate that allows growth of the cells. In Table 4.1 it can be seen that at concentrations of above 10 mM , benzoate has a marked negative effect on the cell densities achieved. On the basis of these cultures a working concentration of 10 mM benzoate was decided upon.

Table 4.1: Shake-flask fermentations to examine effects of benzoate toxicity

Media	Optical Density (after 12 hours)
M9+10g/l glucose	1.8
M9 + 10g/l glucose + 1 mM benzoate	1.509
M9 + 10g/l glucose + 2 mM benzoate	1.497
M9 + 10g/l glucose + 5 mM benzoate	1.052
M9 + 10g/l glucose + 10 mM benzoate	0.873
M9 + 10 mM benzoate	no growth

Initial fermentations run with M9 media and 10 g/l glycerol demonstrated poor growth and a long lag phase. Both of these problems were eliminated by the addition of 0.1% yeast extract as is illustrated in Figure 4.1. This concentration of yeast extract is sufficiently small to have a negligible effect on the CZE analysis (see Figure 3.8).

4.2 Effect of Benzoate and Plasmid Induction

The purpose of the initial fermentations was to determine the effect of 10 mM benzoate and of plasmid induction on the growth of the plasmid-containing organism in a 2 l batch culture. To achieve this, 4 fermentations were carried out, the results of which can be seen in Figures 4.2-4.4. The negative effect caused by the metabolic burden of plasmid expression can be seen from the CER, OUR and dry weight profiles. While this expression is not fatal to the cells it has the effect of reducing the cell growth by over half. In the presence of benzoate as substrate, this effect is counteracted somewhat by the advantage of being able to convert benzoate to pyruvate, which can then enter the central carbon pathway. The toxicity of benzoate is less evident in the induced cells

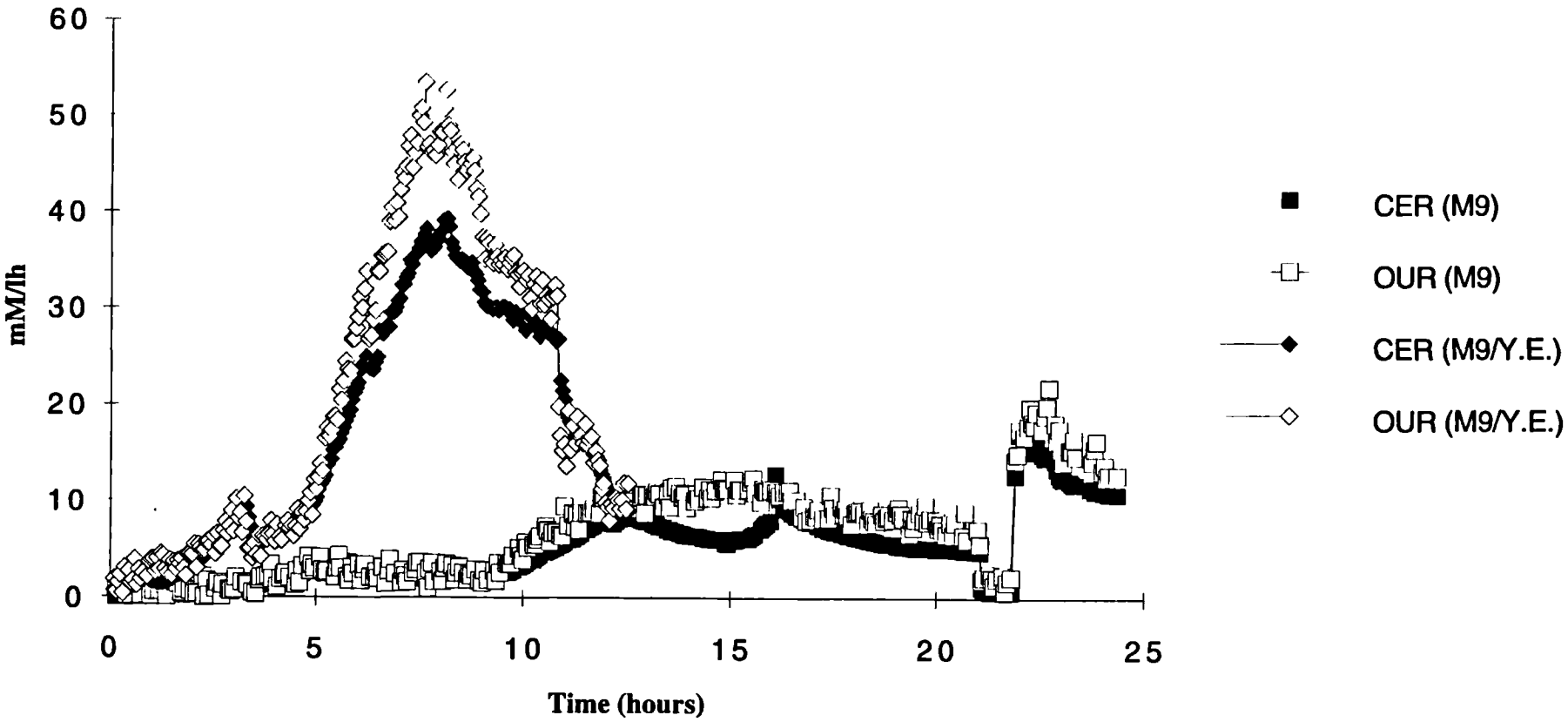


Figure 4.1: This figure shows the effect on the CER and OUR of the strain E.coli JM107 pQR150of adding 0.5% yeast extract to M9 minimal media. The lag time is reduced from 10 hours to 2 hours and the maximum CER and OUR achieved is increased approximately 4-fold. The sharp rise seen in the CER and the OUR of the culture growing on M9 medium at 22 hours was due to the aseptic addition of a small amount of yeast extract to the media.

as it is converted to a carbon source but since the effects are not necessarily additive it is difficult to determine to what extent the contributions to the negative effects on cell growth come from benzoate toxicity, plasmid induction and the toxicity of the pathway intermediates. The RQ profiles shown in Figure 4.4 also vary depending on the substrate and pathway induction conditions, although because of the extent of the noise in the data the trends are less obvious. However it can be seen that the lowest RQ values occur when the pathway is induced and supplied with substrate.

From the above information, the optimum fermentation protocol was defined. The negative effect of plasmid induction upon growth suggested that the cells be exposed to these conditions at the end of the exponential growth phase and just before stationary phase. The toxic effects of benzoate were minimized by ensuring that the pathway was fully induced before its addition.

The same initial fermentations were carried out for strains bearing the plasmids pQR185 and pQR186 (see Figure 3.1 for the pathways encoded by these plasmids). The effect of the substrate p-toluic acid was also examined. This substrate is degraded solely via the dehydrogenase branch of the pathway, whereas benzoate is degraded by both branches (although primarily via the dehydrogenase).

The effects of substrate and plasmid induction on the growth of pQR185 can be seen in Figure 4.5. Induction of the pathway enzymes places a metabolic load on the organism which causes reduced growth rates and a lower final cell density. The addition of the pathway substrate benzoate seems to change this growth pattern, however. Both substrates give higher initial growth rates than are obtained when the pathway is induced in their absence: in the case of benzoate as substrate growth is higher over the first 6 hours and then tails off, for p-toluic acid, the increase in growth occurs between 5 and 13 hours. In both cases, the final cell density is lower that the density achieved with both M9 and M9 with IPTG. This is most likely due to the damaging effect of these substrates on the cell membranes, thus reducing their viability.

The increased initial growth rate effect is also evident from the CER and OUR profiles shown in Figures 4.6 (a) and (b). It can be seen from these that the CER and OUR of *E.coli* pQR185 when the pathway is induced and supplied with benzoate are higher by a factor of two at some points than when no benzoate is supplied and the pathway not induced. These high values for CER and OUR do not occur when p-toluic acid is used

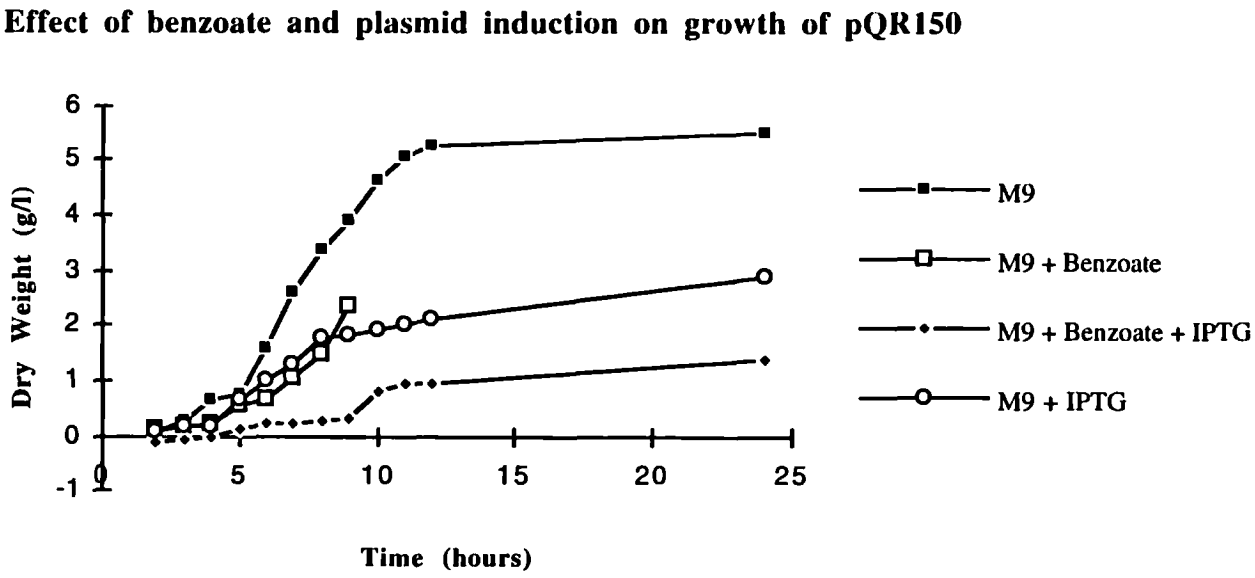
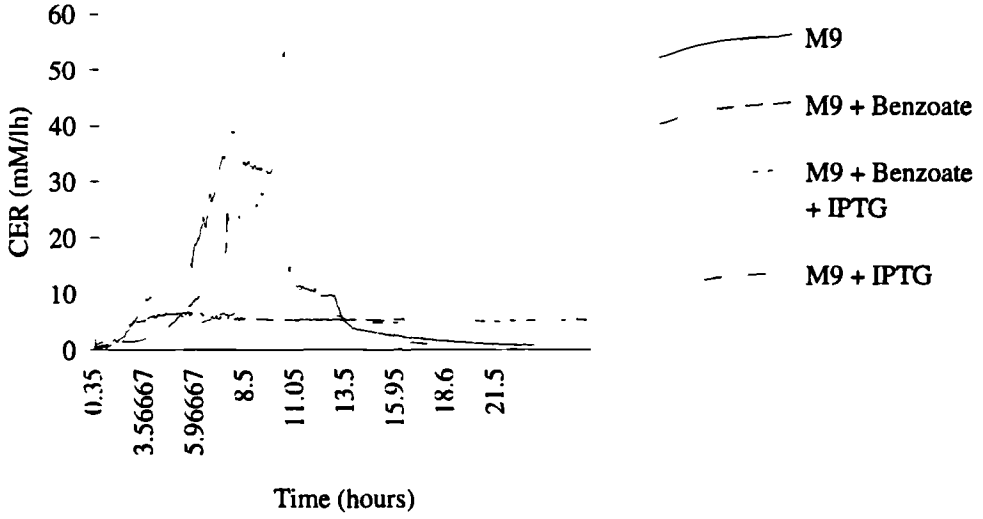


Figure 4.2: This figure shows the effect of plasmid induction and presence of pathway substrate (benzoate at 10 mM) on the growth of E.coli JM107 pQR150. Maximum growth rates and final dry weights of 5.2 g/l were achieved in the absence of the inducer IPTG and the pathway substrate benzoate (■). Addition of 10 mM (❑) benzoate reduced the growth rate and the final cell yield as did the induction of the plasmid by 0.1 mM IPTG (○). The poorest growth was achieved when the pathway was induced and supplied with substrate (◆) where the combined effects of benzoate toxicity and plasmid metabolic load reduce the final biomass concentration to 1.3 g/l.

(a) Effect of benzoate and plasmid induction on CER.



(b) Effect of benzoate and plasmid induction on OUR.

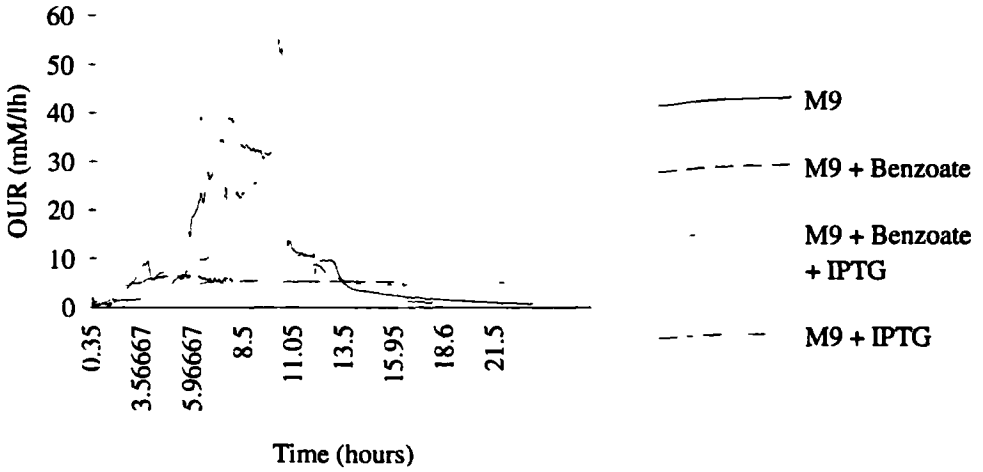


Figure 4.3: This figure shows the variation in the CER (a) and OUR (b) of the strain E.coli JM107 pQR150 with plasmid induction and presence of 10mM benzoate. The effects of benzoate toxicity is not as marked on the CER profile as on the growth profiles .The OUR of the culture whan the pathway is induced and supplied with benzoate is higher that would be expected from the biomass levels because of the oxygen requirements (1 mole of O2 for every mole of benzoate degraded).

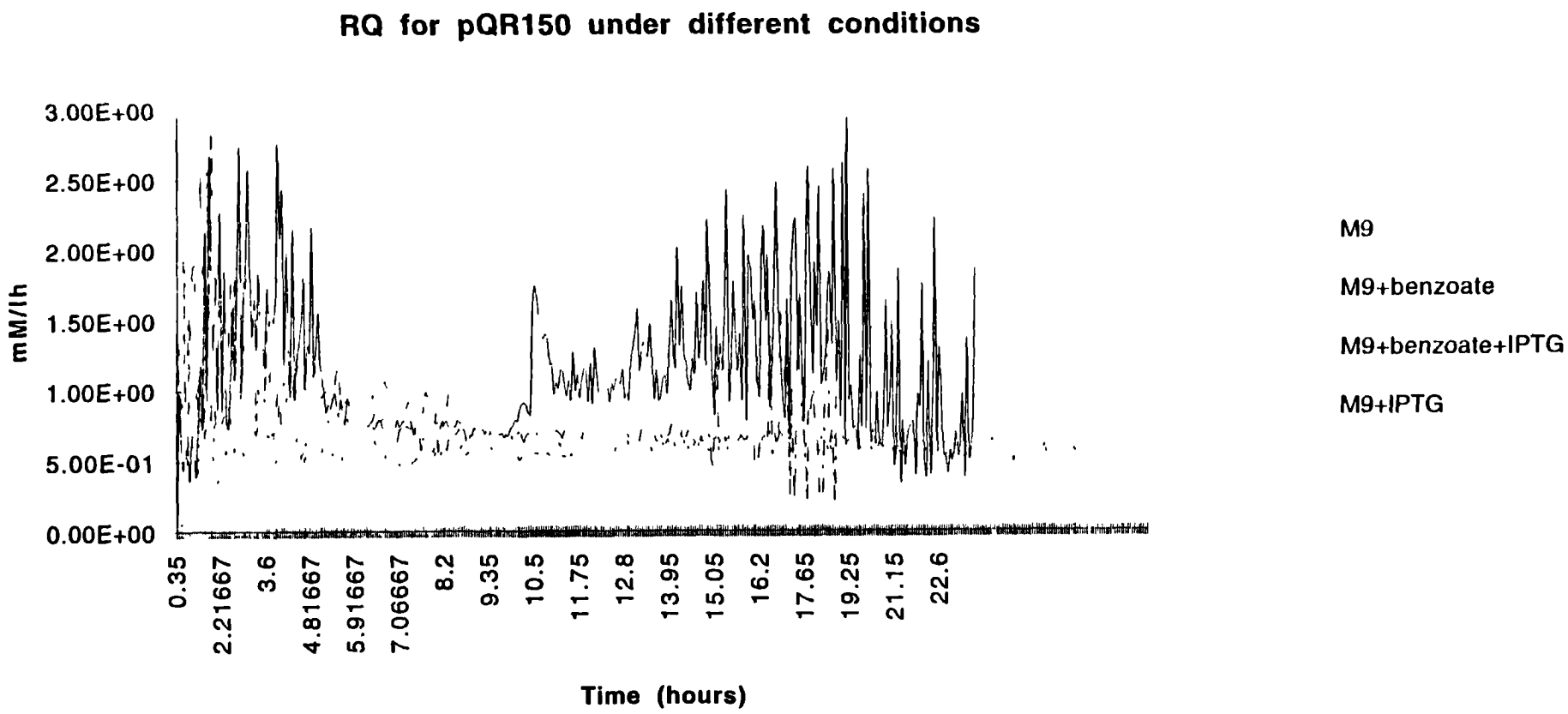


Figure 4.4: This figure shows the lower RQ which occurs as a result of the oxidising activity of the induced pathway when supplied with benzoate. It also illustrates how the noise in the CER and OUR data is much more significant when looking at the quotient of the two.

as a pathway substrate. In this case, values for CER and OUR are lower than those obtained when no substrate is supplied. Values of CER and OUR are lowest for the strain when the pathway is induced but no substrate is supplied.

Looking at the effect of fermentation conditions on the growth of pQR186 in Figure 4.7, similar trends are evident. The presence of benzoate at a concentration of 10 mM when the pathway is not induced causes a higher growth rate between 7 and 12 hours but a lower final cell density than when benzoate is absent. Induction of the pathway enzymes in the presence of benzoate has a further negative effect on cell growth and final cell density. This is due to both the metabolic load placed on the cell by the expression of the pathway enzymes and the toxic effects of the pathway intermediates. The use of p-toluic acid as pathway substrate has a much more damaging effect on the cell. There is almost no growth until after 14 hours and the final cell density is significantly lower than each of the other cases. The CER and OUR profiles in Figures 4.8 (a) and (b) correspond well with the growth trends. All of the CER and OUR traces show the same pattern of rising gradually to a highest point and then falling off suddenly. What differs is the length of time over which this occurs and the maximum CER or OUR value achieved. The greatest CER and OUR values (31.5 and 45.2 mM/lh respectively) occurred during the growth of E.coli pQR186 on benzoate without the pathway being induced. After a 6 hour lag phase there was an increase over the following 9 hours until the values dropped sharply. The time course was similar when the pathway was induced and supplied with benzoate but the CER and OUR values were lower, the maxima being 14.8 and 20.2 mM/lh respectively. When p-toluic acid is supplied as a substrate to the induced pathway, the CER and OUR begins to rise only after 15 hours, reaching maxima of 27.5 and 35.1 mM/lh respectively at 28.5 hours and then dropping off. The lowest CER and OUR values correspond with the highest final cell density achieved and occur when growing the strain on M9 without inducing the pathway or supplying it with substrate.

There is a lot of noise in the on-line fermentation and gas analysis data as can be seen in the CER and OUR profiles shown in Figures 4.2 to 4.8. The existence of out-lying points in the gas analysis data was also a problem.

There were several causes for these outlying points in the data and these varied depending on whether the data acquisition system was Bio-i or RT-DAS. The first was

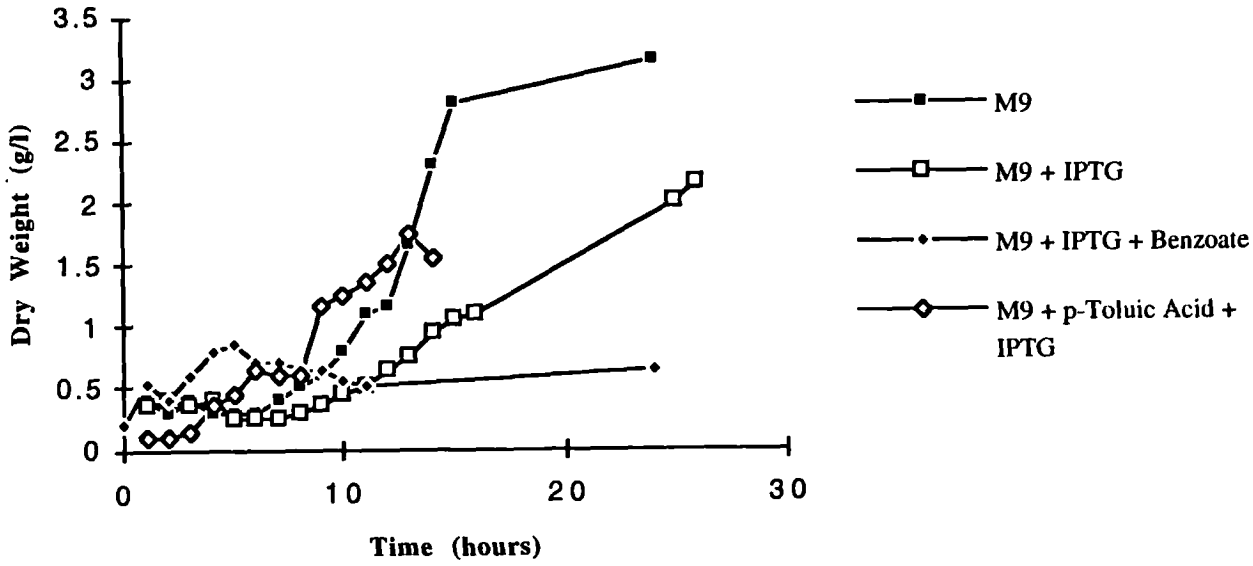
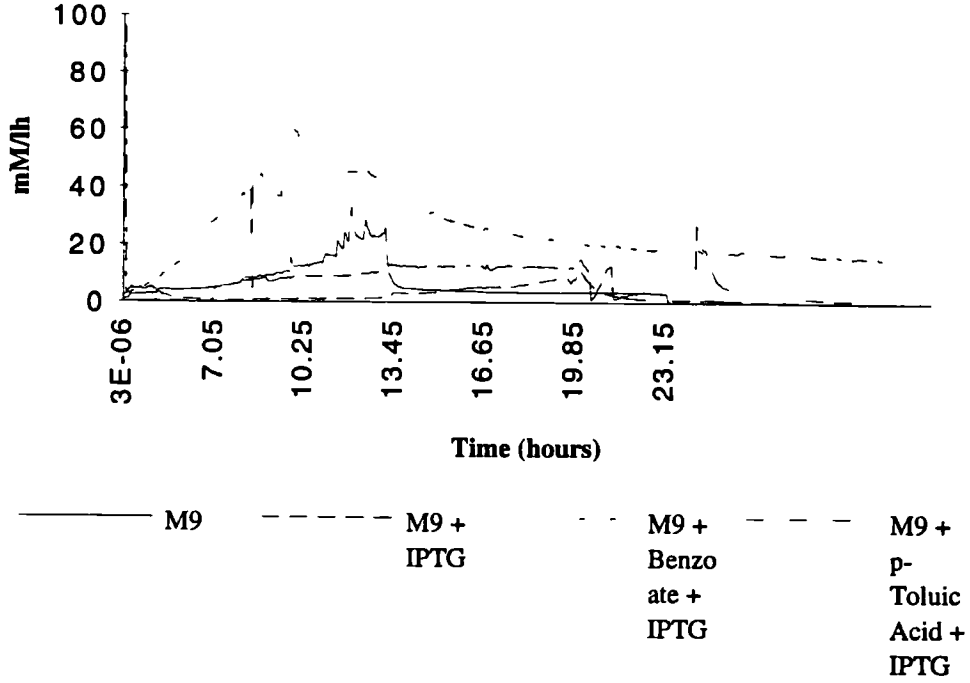


Figure 4.5: Effect of substrate and plasmid induction of growth of E.coli JM107 pQR185.
the initial growth rate of the strain with benzoate and IPTG is greater than for any of the other cases, but the final yield is lowest. Again the strain grows best without plasmid induction or pathway substrate. The best biomass yield is a little over half that achieved with the plasmid pQR150. This is attributable to both the build-up of pathway intermediates due to the pathway truncation and the fact that pyruvate is not being supplied to the central carbon pathway.

pQR185



(b) Effect of substrate and plasmid induction on OUR of pQR185

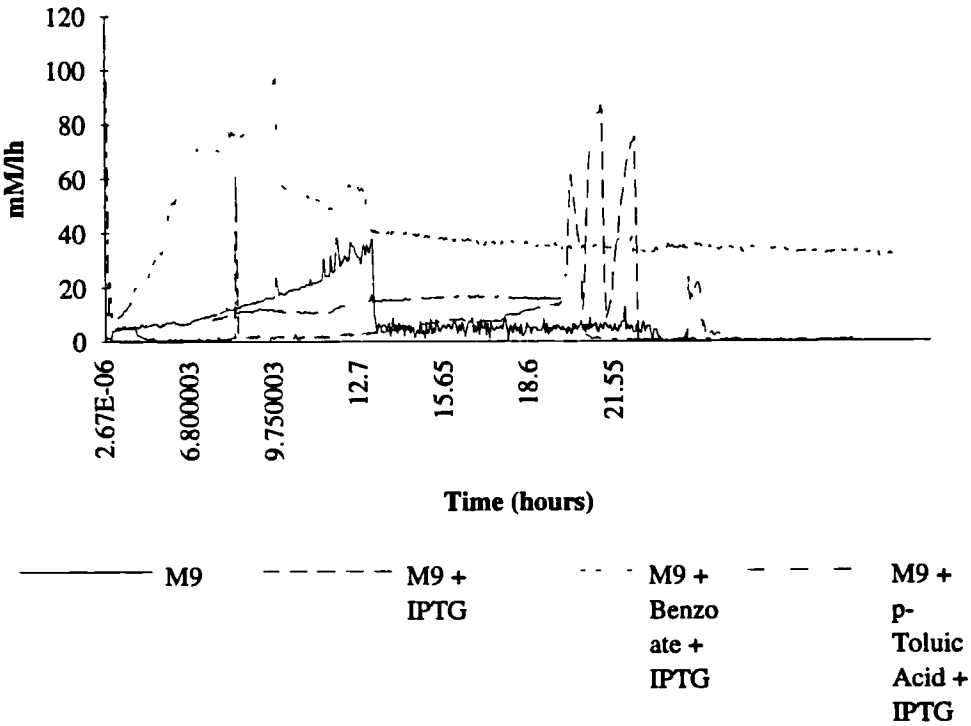


Figure 4.6: Variation of (a) CER and (b) OUR for the strain *E.coli* JM107 pQR185 resulting from presence of pathway substrate and pathway induction by IPTG.

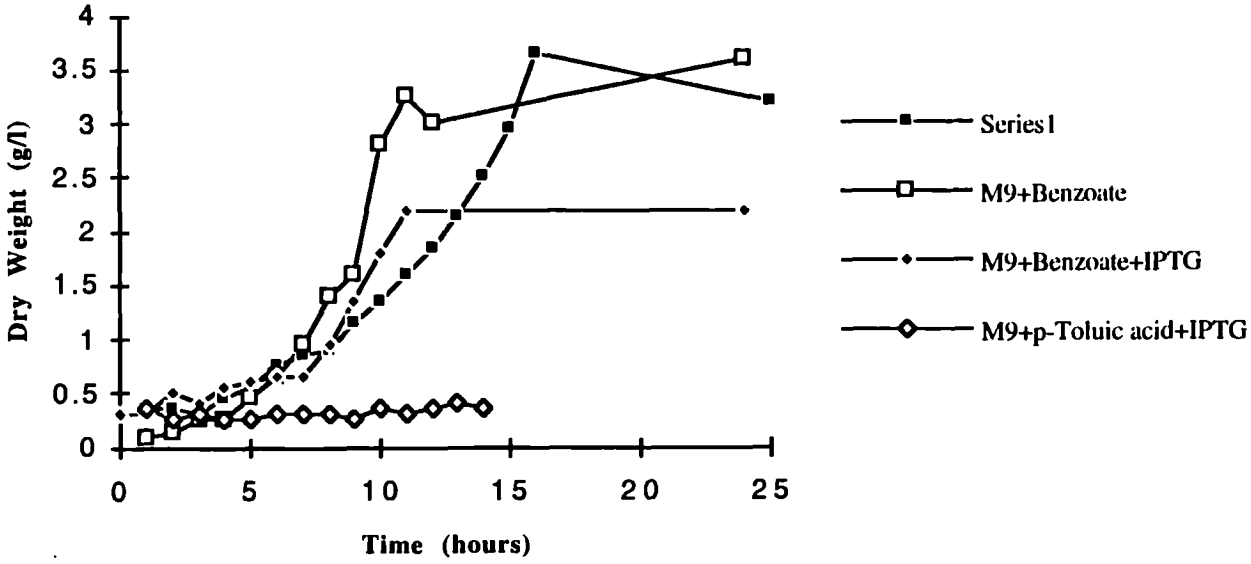
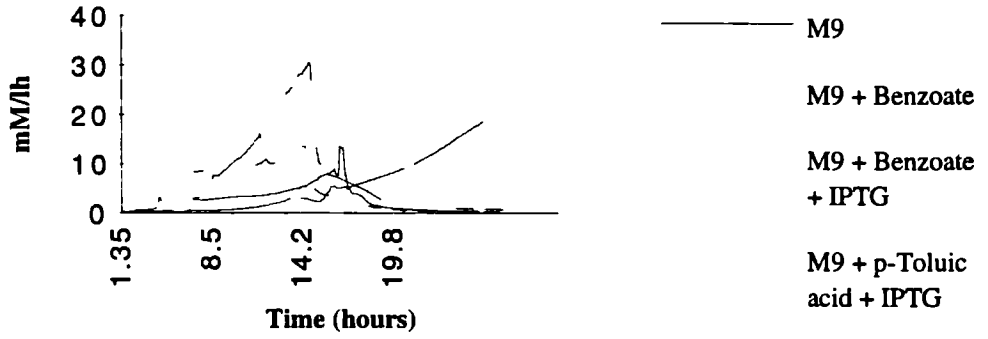


Figure 4.7: Effect of substrate and plasmid induction on growth of pQR186
For this plasmid, all of the initial growth rates are similar. The highest cell density is achieved during growth on M9 only. As with pQR185` the maximum cell density achieved is just over half that obtained for the full pathway. The inhibitory effect of benzoate is less noticable in this case but it can be seen that p-toluic acid has a severely inhibitory effect on this strain.

(a) Effect of substrate and plasmid induction on CER of pQR186



Effect of substrate and plasmid induction on OUR of pQR186

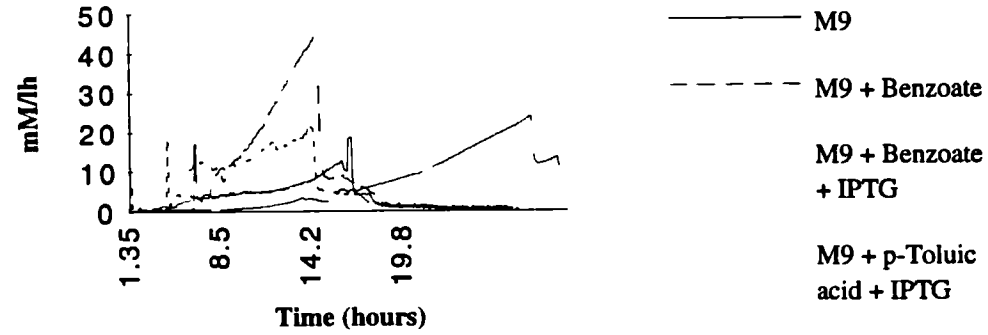


Figure 4.8: Variation of (a) CER and (b) OUR for the strain *E.coli* JM107 pQR186 resulting from presence of pathway substrate and pathway induction by IPTG.

the interface between the mass spectrophotometer and the data acquisition system which returned sporadic values of -9999 due to signal interruption. These occurred irrespective of what data acquisition system was used. The second type of error was in the positioning of the decimal point which occurred when using RT-DAS. An example is shown in Table 4.2. The out-liers generated by this kind of error were much more difficult to identify and eradicate than the first type, which were dealt with by a writing an Excel macro. Setting an upper limit will eradicate some erroneous points but it is generally not possible to set a lower limit. Also, as with any computer-based data acquisition system there were a number of crashes during the fermentations and these left "holes" in the profile data. While these can in most cases be easily removed when analysing data historically, they can be a serious problem if data is to be used in on-line control or decision-making and must be recognised if spurious results are to be avoided.

Table 4.2: Off-gas analysis data, showing the different errors which can occur (underlined)

Time (hours)	CER (mM/lh)	OUR (mM/lh)	RQ
4.700287	43.014	76.537	0.562
4.840843	0	0	0
4.940843	40.601	75.535	0.5375
4.990843	40.375	75.831	0.5324
5.040843	-9999	-9999	1
5.090843	46.311	82.103	0.564
5.140843	46.224	82.709	0.5588
5.190843	46.2	82.361	0.5609
5.240843	37.035	773.51	0.0478
5.290843	36.657	765.76	0.0478
5.340843	36.657	765.76	0.0478
5.390843	36.657	765.76	0.0478
5.440843	46.704	76.842	0.6077

The presence of questionable measurements in the data can be seen in the RQ profiles in Figure 4.4. Because the RQ is a quotient of the CER and the OUR it is very sensitive and any fluctuations in these data sets are magnified. Because of this it is difficult to identify trends.

4.3 Optical Density and Dry Weight Correlations

Throughout the range of the fermentations on pQR150, pQR185 and pQR186, optical density (O.D.) was a reliable indicator of biomass, even though the yellow compound hydroxy-muconic semialdehyde was produced in fermentations where the pathway was expressed. The correlation is shown in Figure 4.9. This shows the optical density as a function of the dry weight obtained throughout the time courses of 15 fermentations. This gives a linear fit with a coefficient of determination (R^2) of 0.91. The different symbol sets in this figure represent data from different fermentations. It can be seen that the different data sets have slightly different correlations. The result of this is that the individual data sets provide a better correlation (with an R^2 of approximately 0.97) for a particular set of fermentation conditions than does the overall correlation, which should be used in the absence of an individual correlation.

Table 4.3: Elemental analysis of *E.coli*

Element	% of dry weight	Element	% of dry weight
Carbon	50	Sodium	1
Oxygen	20	Calcium	0.5
Nitrogen	14	Magnesium	0.5
Hydrogen	8	Chlorine	0.5
Phosphorus	3	Iron	0.2
Sulfur	1	All others	0.3 approx.
Potassium	1		

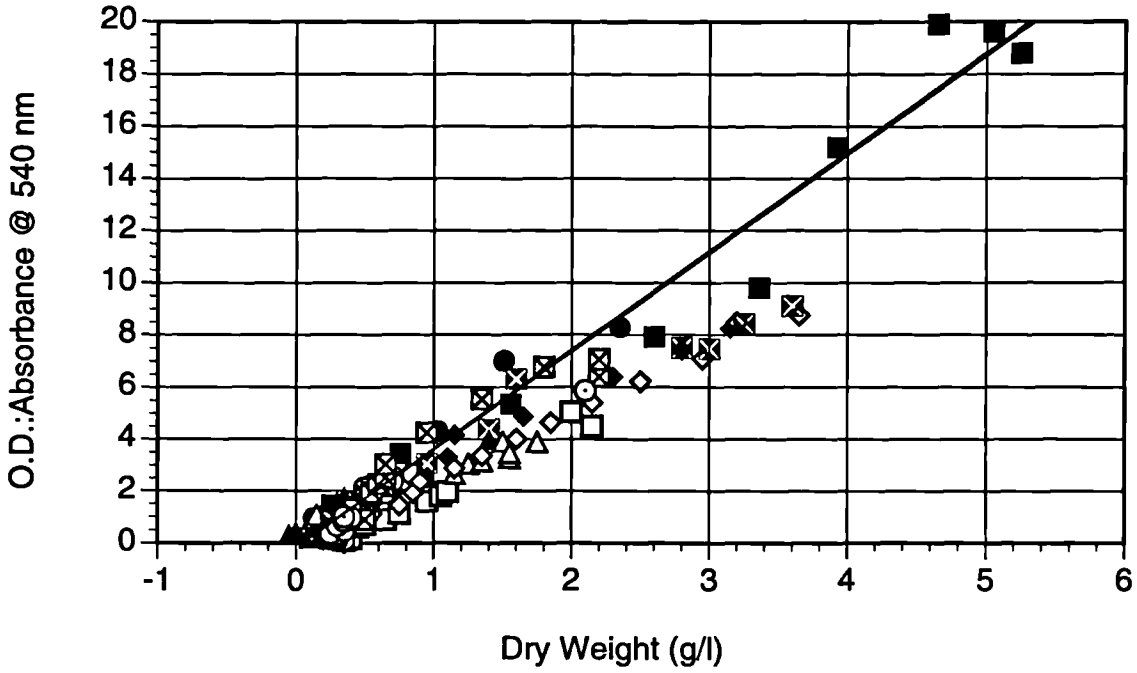


Figure 4.9: Correlation of dry weight and optical density

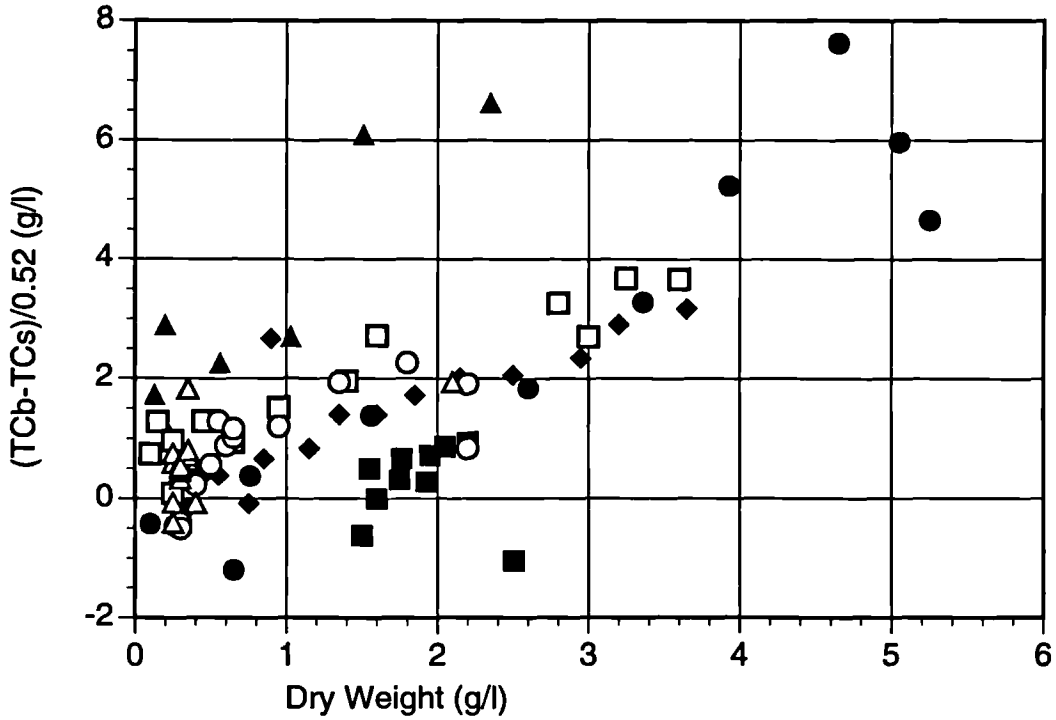


Figure 4.10: Correlation of biomass measured by dry weight with that measured by total carbon

This indicates that this correlation can be used to provide a rapid estimate of biomass for any fermentation of *E.coli* JM107 on M9 irrespective of any plasmid that it may be carrying and whether or not the pathway is expressed.

4.4 Results of Total Organic Carbon Analysis

Total organic carbon analysis was performed on the whole broth and on the supernatant from centrifuged samples. Total organic carbon is obtained by measuring the total carbon and the total inorganic carbon and subtracting the two. The difference in carbon content between the broth and the supernatant has previously been used as an estimator of the biomass. An attempt was made to correlate the dry weight measured by carbon analysis and that measured by spinning down, drying and weighing cells in eppendorfs. To do this, the elemental analysis of *E.coli* shown in Table 4.3 (Bailey and Ollis, 1986) which gives the percentage of carbon present in an E.coli cell was used together with the difference between the total carbon in the broth and the supernatant to give the biomass.

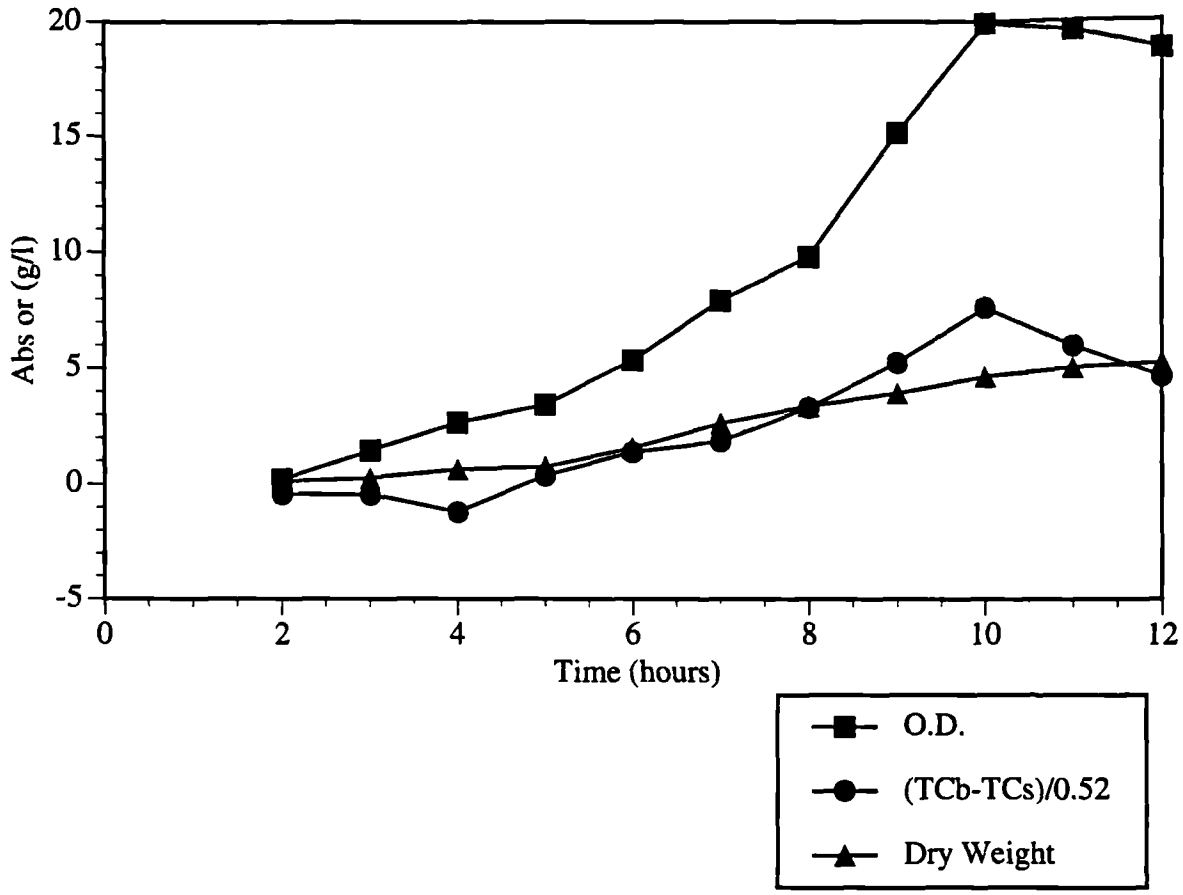
Plotting the total carbon in the biomass ($biomass_{TC}$) against the dry weight over 15 fermentations (see Figure 4.10), we find a much poorer correlation that obtained for the optical density. A linear regression gives an R^2 value of 0.001. The scatter obtained in these measurements was not reflected in repeatability of analysis of individual samples, all of which showed a coefficient of variance of less than 2% over three measurements. The repeatability was not noticeably worse in outlying points than in points which gave the predicted result so that it was unclear whether these deviations could be attributed to measurement error or whether they were reflections of what was happening in the culture. These total carbon measurements are later used in carbon balancing and can be used in on-line flux analysis. For this reason it is crucial to be able to decide whether they are accurate measurements of the state of the culture, and when a data point falls out of the region which is physiologically reasonable. To isolate possible reasons for this variance and more closely identify its occurrence, the total carbon data was looked at in more detail.

A problem with analysis of the errors in this data is that there is no perfectly correct set of data with which to compare our results, so that the scatter in Figure 4.10 may

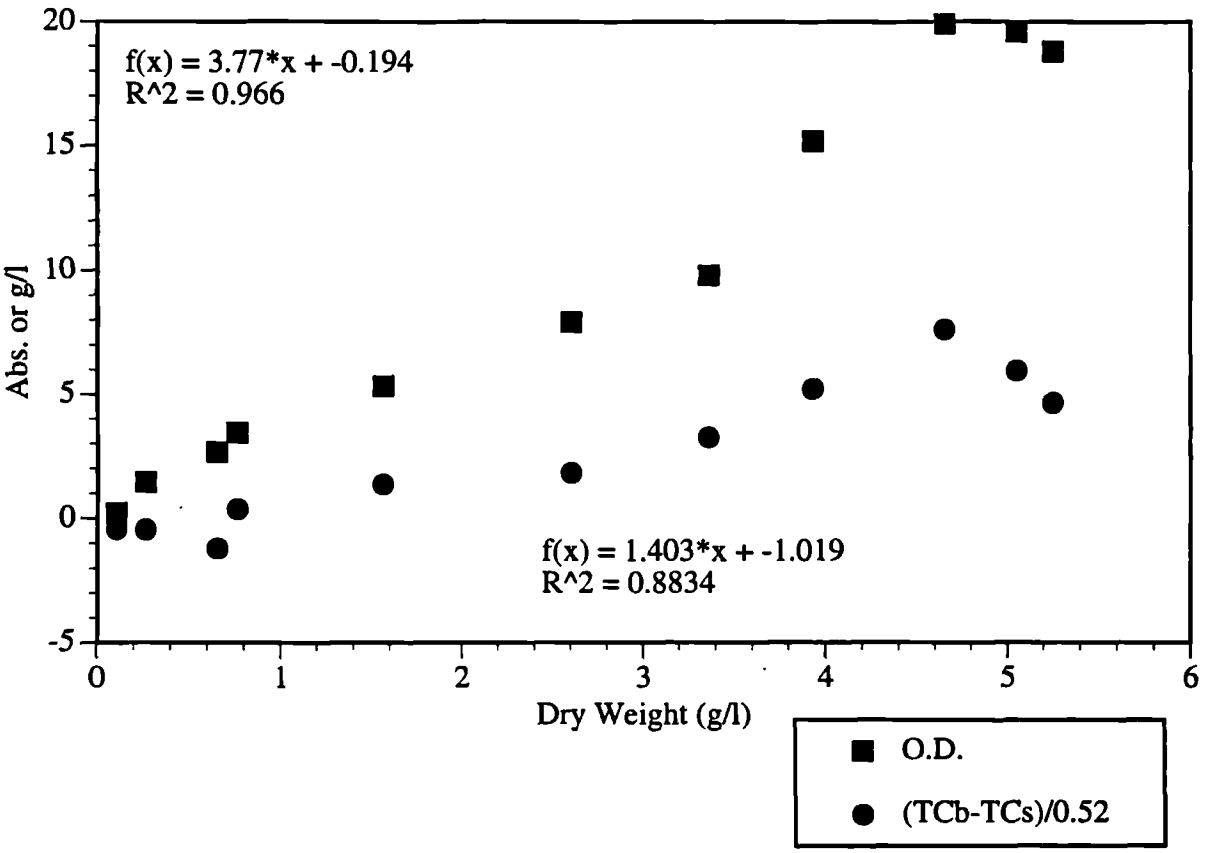
be due to errors in the dry weight data as well as any present in the total carbon data. For this reason it is instructive to use the dry weight:optical density correlations and the Carbon emitted as CO_2 (C_{CO2}):Total Carbon in broth (TC_{broth}) correlations to aid in identifying where the error lies. For example, Figure 4.11a shows the profiles of the optical density, the biomass as measured by dry weight (d.w.) and the biomass as measured by total carbon ($biomass_{TC}$) over the time course of a fermentation of *E.coli* JM107 pQR150 growing on M9/0.1% yeast extract media. It is seen that after 10 hours there is a peak in the O.D. and $biomass_{TC}$ profiles which does not occur in the d.w. profile. This is reflected in how these measurements correlate with each other in Figure 4.11b such that O.D.:$biomass_{TC}$ gives a better correlation as reflected in the coefficient of determination, R^2, than does d.w.:$biomass_{TC}$. This indicates that some of the error may be in the dry weight measurements rather than in the total carbon measurements. However, looking at how the carbon emitted as CO_2 corresponds with the decrease in total carbon in the broth in Figure 4.11c, it can be seen that there is an outlying point which corresponds to the 10 hour data in Figure 4.11a, signifying that the principal proportion of the error resides in the total carbon data. If this data point were to be considered as an outlier, then the R^2 for the d.w.:$biomass_{TC}$ improves to 0.966 and the C_{CO2}:TC_{broth} changes to 0.981.

One feature of the biomass as predicted by total carbon measurements is the notable fluctuations in the data in the early stages of the fermentation, up to mid-log phase. Because the biomass values being measured here are low, any experimental errors such as those from dilution and the retention of media components by the biomass to be weighed will be disproportionately reflected. This can be seen in Figures 4.12 and 4.13 which show the growth of strains containing the plasmids pQR185 and pQR186 growing on M9/0.1% yeast extract media. In Figure 4.12, while there is a lot of scatter initially, this ceases once the biomass reaches 1 g/l. A correlation using all data points gives an R^2 of 0.84 whereas if only those data points over 0.75 g/l d.w. improves the correlation slightly giving an R^2 of 0.88 but the slope is not greatly affected, changing from 1.03 to 1.05.

In Figure 4.13, it can be seen that the $biomass_{TC}$ data corresponds with the biomassd.w. quite well except for one outlier at 8 hours. This outlier results from the TC_{broth} data and is also present when comparing the carbon emitted as CO_2 with the total carbon in



gure 4.11 (a): The time profiles of the three means of measuring biomass, dry weight, optical density nd total carbon measurements



e 4.11 (b): This figure shows the correlations between dry weight and optical density and between dry ıt and biomass as measured by total carbon. The best fit lines and correlation coefficients are also n.

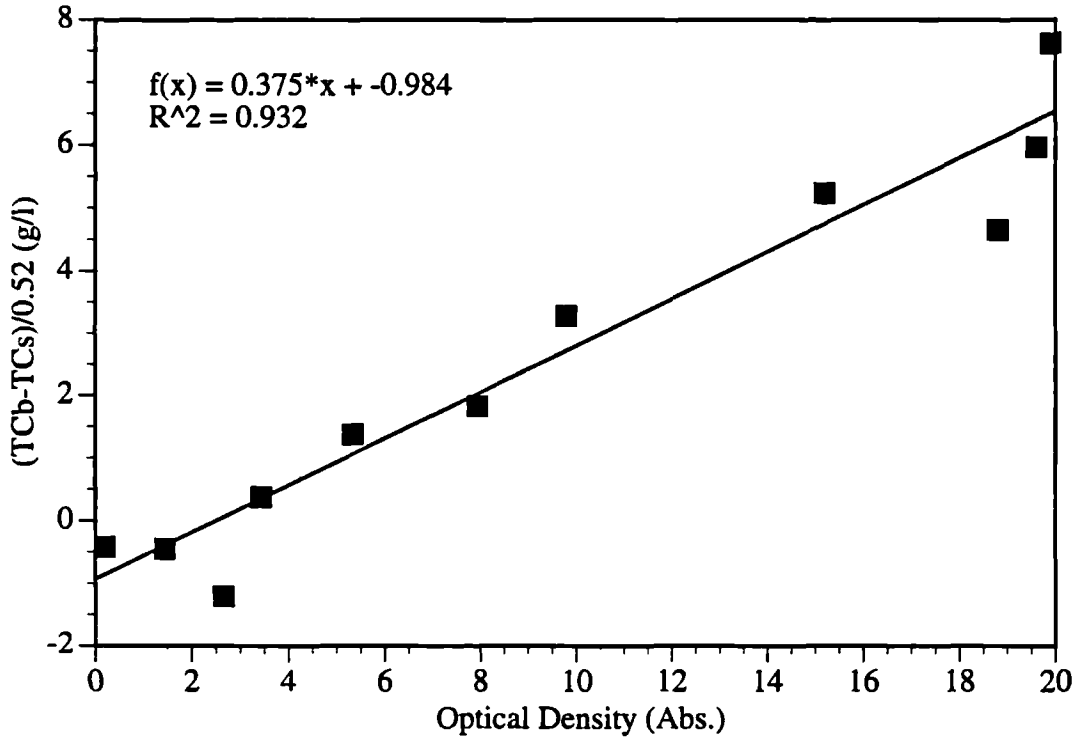


Figure 4.11 (c): This figure shows the correlation between optical density and bioomass as measured by total carbon. This correlation gives a better linear fit than those in (b) as can be seen in the R^2 values.

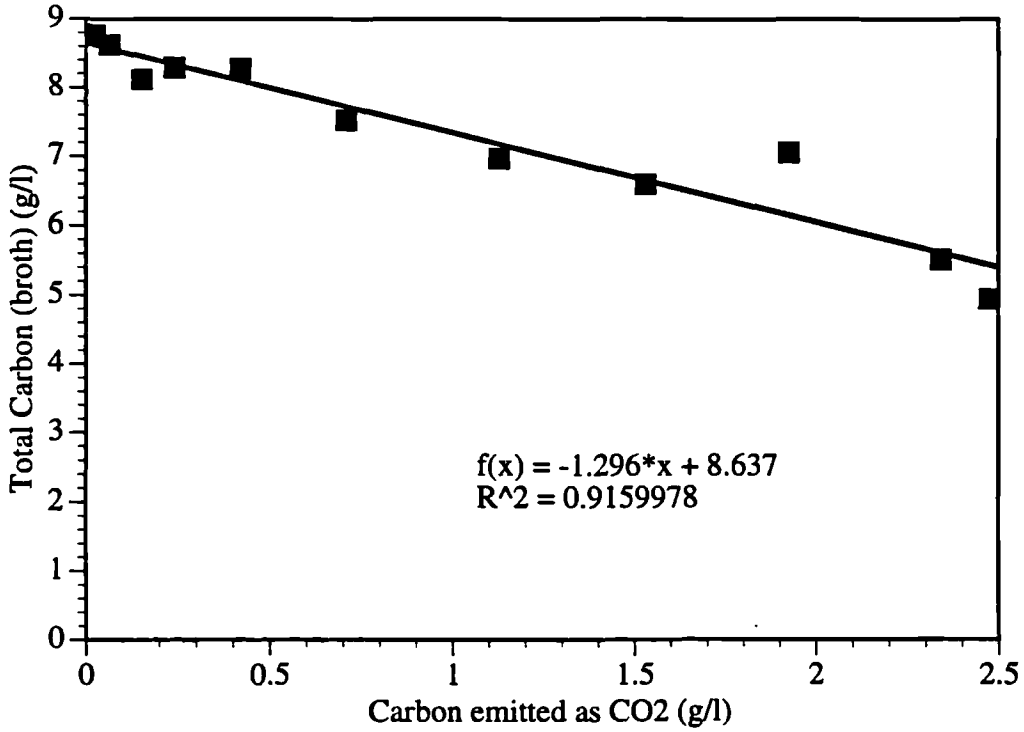


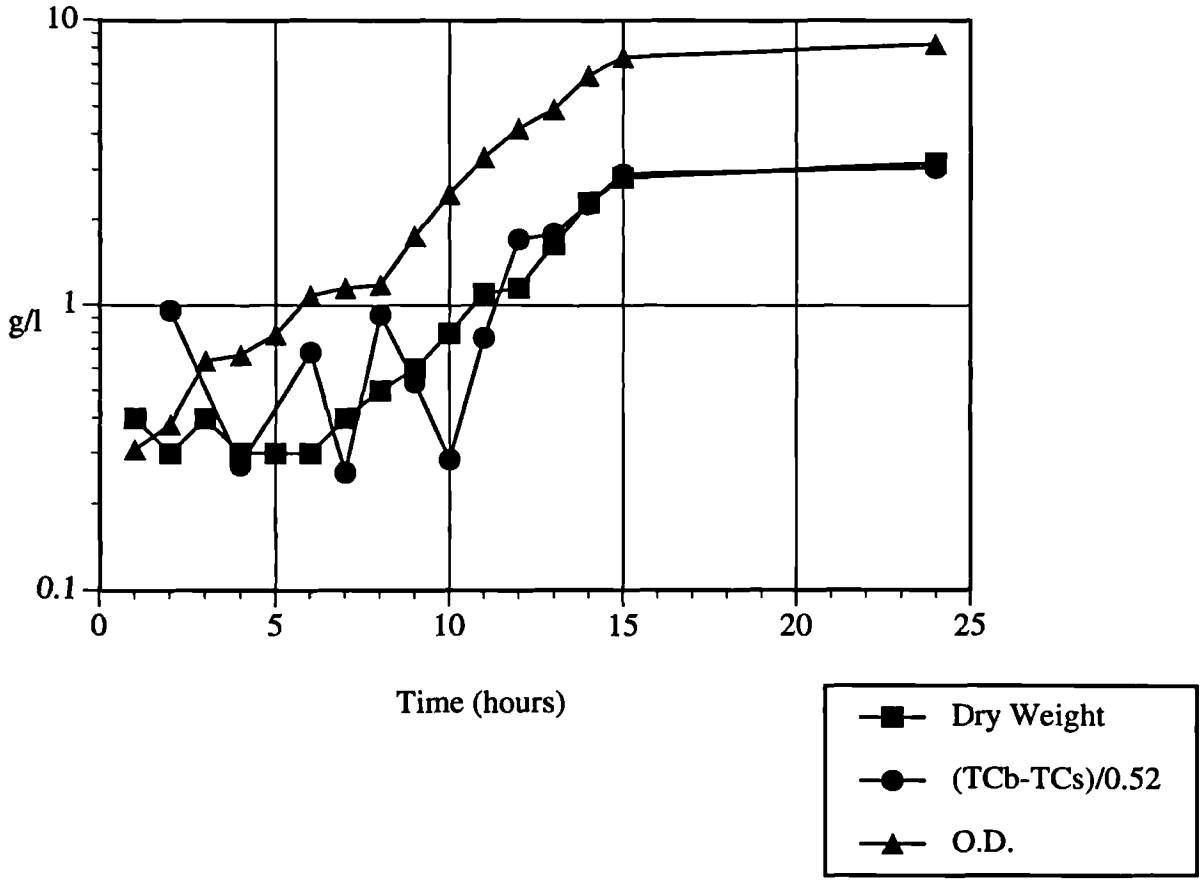
Figure 4.11 (d): This figure shows the correlation between total carbon in the broth and the carbon emitted as CO2.

the broth. If this point is regarded as lying outside the confidence limits of the data, the R^2 value for the d.w.:$biomass_{TC}$ correlation improve from 0.75 to 0.94 and the R^2 value for C_{CO2}:$C_{TCbroth}$ becomes 0.96.

It should be noted that in this case and in other cases where outliers occurred in the $biomass_{TC}$ data, the variance was in the measurements of total carbon in the broth and not in those of total carbon in the supernatant. This suggests that continuing metabolic activity which may have occurred during handling and analysis of the sample may be one reason for spurious results.

Despite the variability in the d.w.:$biomass_{TC}$ correlations, the C_{CO2}:TC_{broth} correlation was, over all of the fermentations considered, quite good with a minimum R^2 of 0.80. The carbon emitted as CO_2 is calculated from the cumulative CER using numerical integration. Because the data is sampled every 3 minutes, the errors introduced by this method are small. However, hardware or software malfunctions may have a significant effect depending on the CER levels being measured. The slopes of the regression lines are in general greater than one, indicating that there is a difference in the calibration of the two measurement systems, the mass spectrometer pumps and the total carbon calibration.

As mentioned above, further metabolism during sample preparation and analysis can introduce errors not only because the production of CO_2 and its subsequent sparging will cause an under-estimation of the total carbon present in the culture but also further metabolism will effect a change in the pH of the sample and this will alter the equilibrium of CO_2 in the sample. Another source of error could arise from inadequate or variable purging and this will result in an over-estimation of the total carbon present. The total carbon data can be used along with the CER data to perform a carbon balance over the system. This carbon balance can be used to provide estimates of the carbon flux to by-products and of the carbon source concentrations in the fermentation. This method and the results from the fermentations described above are outlined and discussed in Section 5.1.1.



re 4.12 (a): Profile of dry weight, biomass(TC) and O.D. with time during growth of *E.coli* pQR185 on M9/Y.E.

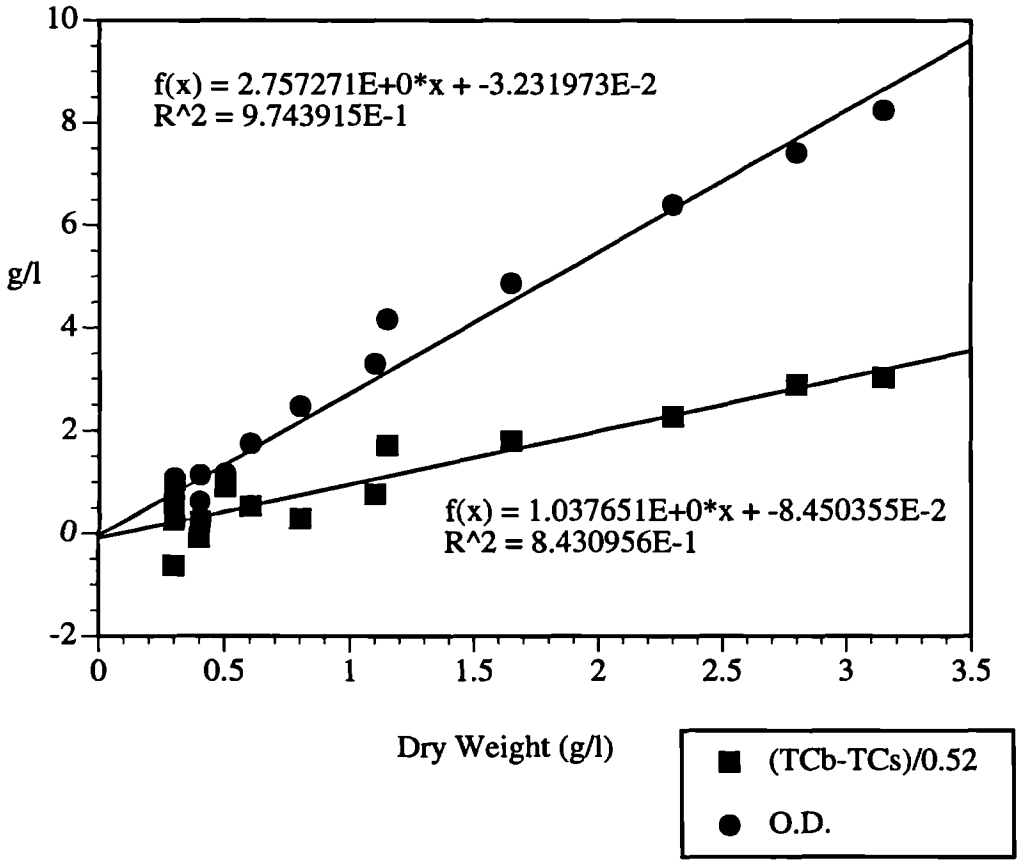


Figure 4.12 (b): Correlation of dry weight with biomass(TC) and O.D.

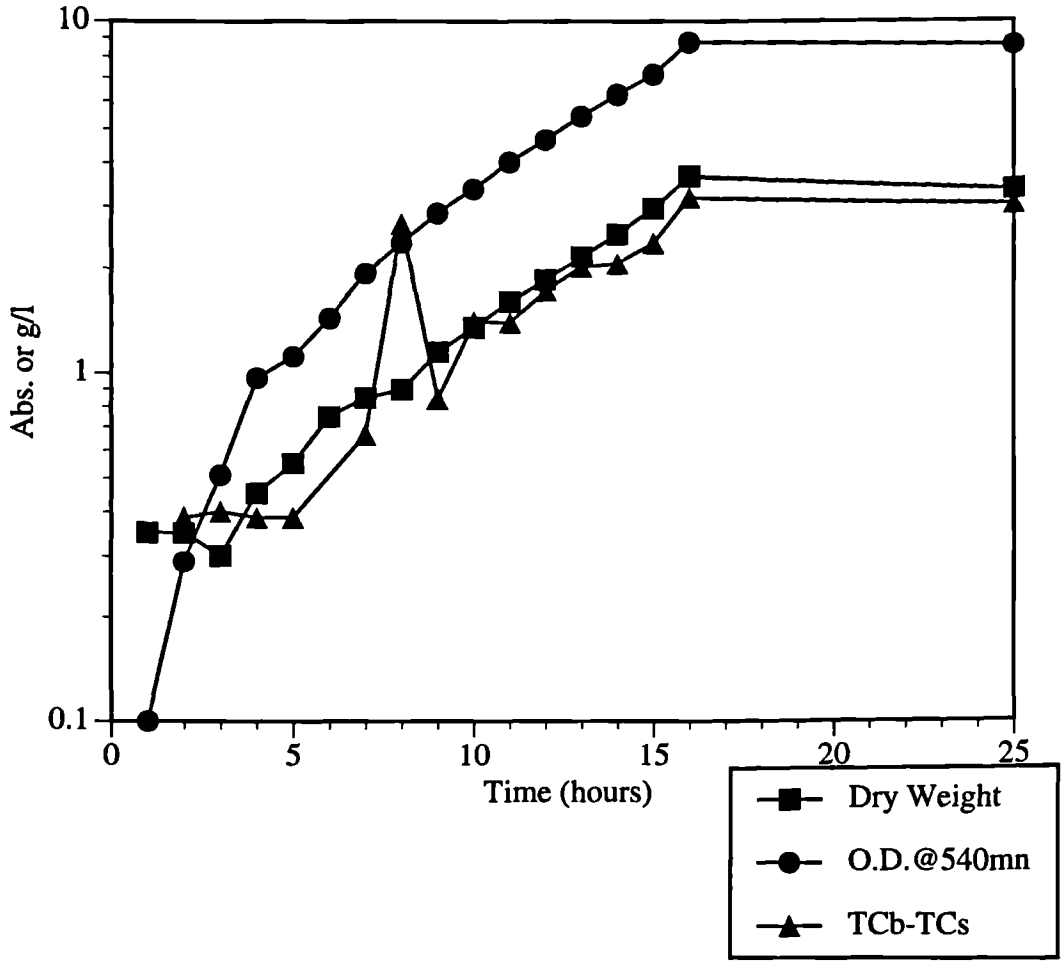


Figure 4.13 (a): Time profiles of dry weight, optical density and total carbon measurements during growth of E.coli JM107 pQR186 on M9 medium.

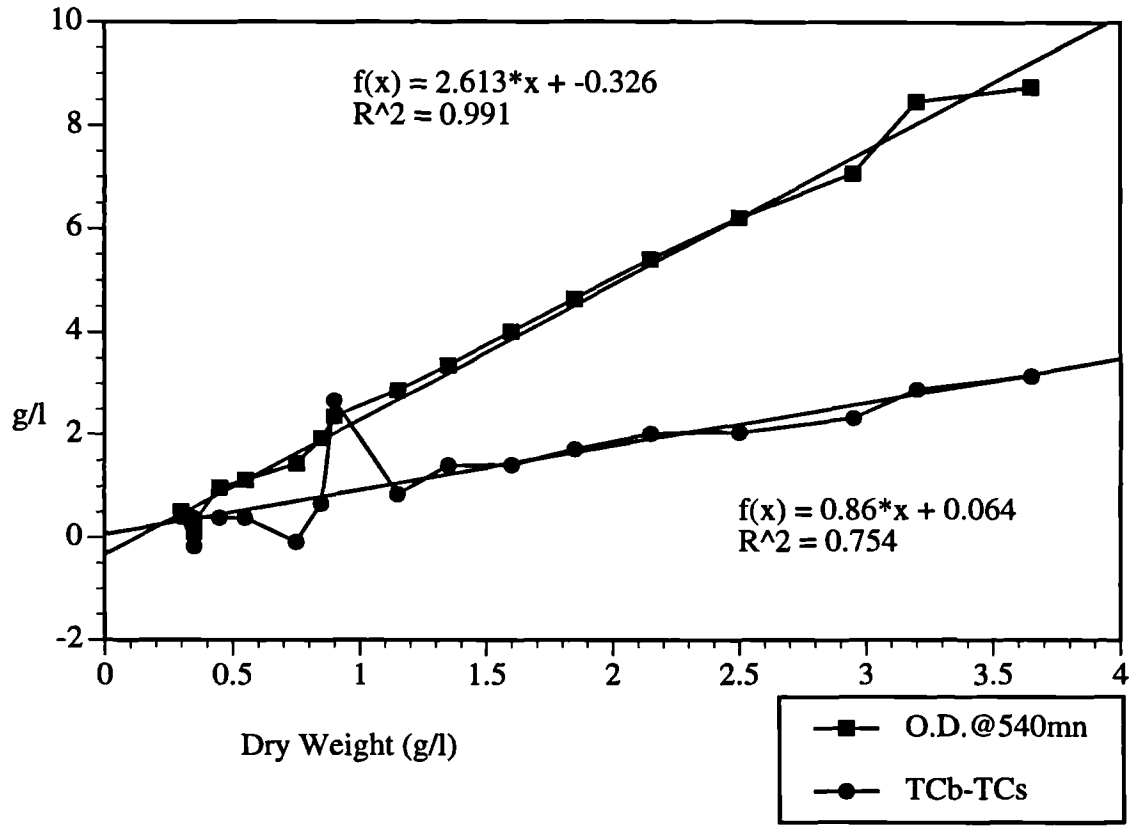


Figure 4.13 (b): Correlation of dry weight with optical density and with biomass as measured by total carbon.

4.4 Total Nitrogen Data

Total nitrogen data was collected to monitor the nitrogen consumption of the culture over the course of the fermentation. It is theoretically possible to use this as an indicator of dry weight, but the accuracy is significantly worse than that obtained by total carbon measurements or by weighing. Also, because of the low levels of nitrogen present in the fermentation broths studied (typically 0.01 %), it is necessary to concentrate samples by evaporation in order to obtain reasonable results. This also removes any time advantage as the analysis is the same as that for centrifuging, drying and weighing cells.

Total nitrogen measurements can however be used as a measure of the total protein content of a stationary culture. This data can be used to monitor protein production as in, for example, the expression of the enzymes for a pathway.

4.5 Metabolite Concentrations

Measurements of specific metabolites were carried out by several means. Glycerol was measured by commercial enzyme assay kits. Pyruvate, acetaldehyde and formate were assayed by CZE with back-up analysis carried out by enzymic assay kits. Benzoate, catechol, dienoate and oxalocrotonate were measured by CZE and hydroxymuconic semialdehyde was assayed spectrophometrically. All of the above methods are described in Section 3.4.

Analyses were carried out on the broth supernatant and also on the sonicant supernatant in order to detect differences between intra- and extra-cellular levels of the metabolites. Figure 4.14 shows the measured levels of hydroxymuconic semialdehyde from the broth supernatant and the sonicant supernatant. It can be seen that there is no significant difference in the levels except at the 24-hour point where concentrations are higher in the broth supernatant than in the sonicant supernatant. This is most probably caused by the absorbance of the compounds onto the products of cell lysis which will be much more prevalent after 24 hours than during the growth stage of the fermentation. These results (which were similar for the other intermediates measured) indicate that the

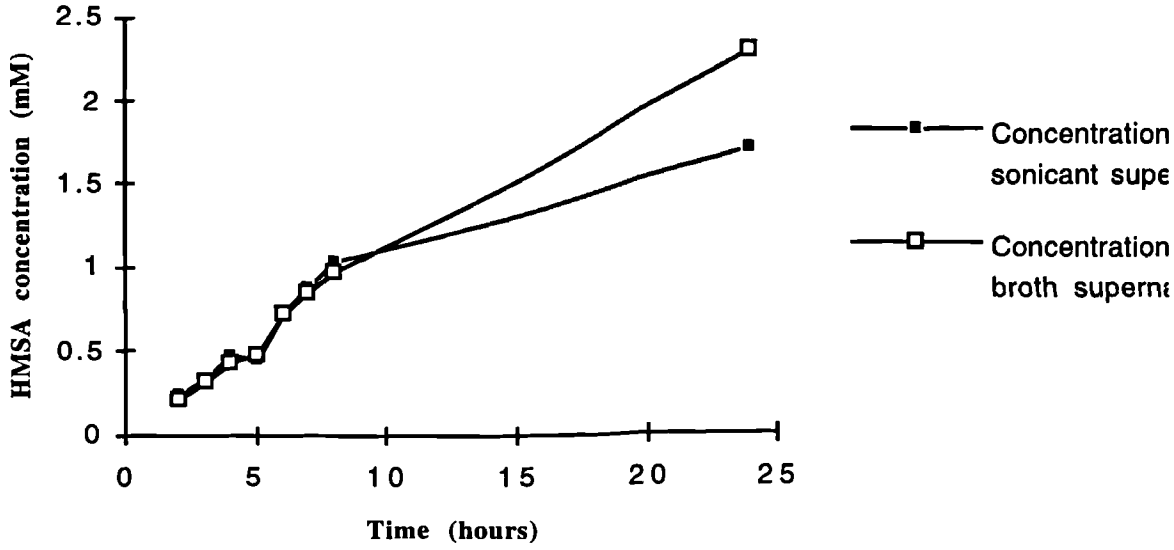


Figure 4.14: This figure shows the concentrations of hydroxymuconic semialdehyde as measured in the broth supernatent and in the sonicated supernatent. The similarity of the results indicates that there is no signifigant difference between intra- and extra-cellular levels of the metabolite.

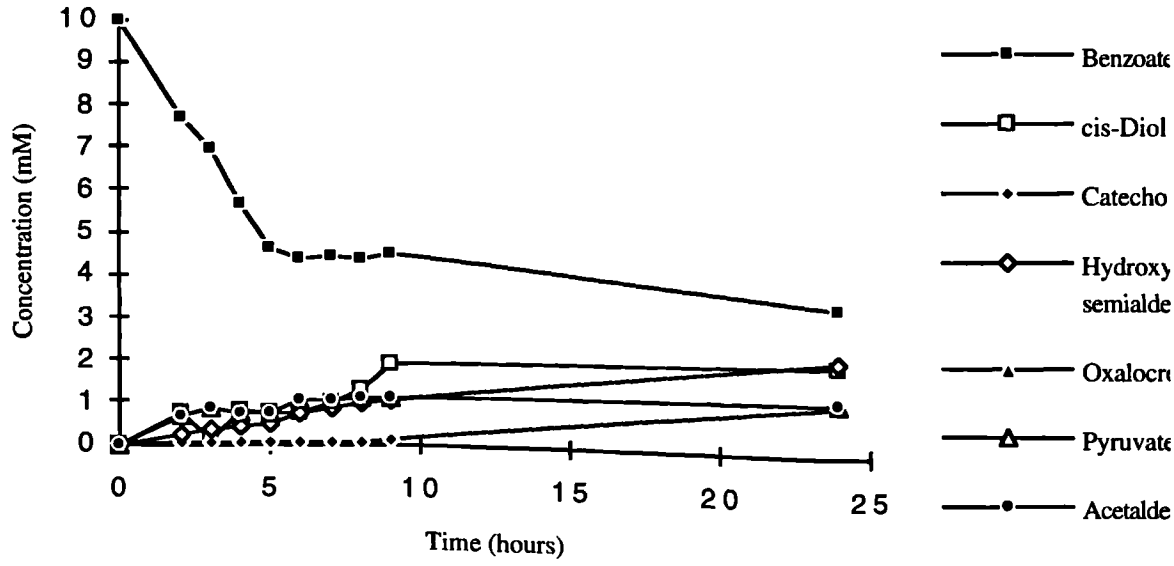


Figure 4.15: This figure shows the concentrations of the pathway intermediates over the time course of the stationary phase of a fermentation.

extracellular intermediate concentrations can be used as a good approximation of the intracellular intermediate concentrations. The reason this leakage of the pathway intermediates occurs may be due to the damaged cell membrane caused by the toxic effects of benzoate. E.coli may also actively secrete the pathway intermediates to minimise their toxic effect.

Because *E.coli* JM107 has no pathways for aromatic degradation, the pathway intermediates were only measured for those fermentations where pathway substrate was added and the pathway induced. The profiles of the intermediates benzoate, cis-diol, catechol, hydroxymuconic semialdehyde and oxalocrotonate are given in Figure 4.15. those for pyruvate, formate and acetaldehyde are shown in Figure 4.16. The profiles in Figure 4.15 show the benzoate concentration decreasing at first rapidly, then levelling off to a concentration of 2.5 mmol. This effect may be attributed to a decrease in enzyme viability or perhaps to deactivation of the pathway enzymes by feed-back repression of intermediates. Similarly the other pathway intermediates build-up quickly at first and then level off to a steady concentration.

It is these concentration profiles which will be used to calculate the control coefficients of the pathway in Section 5.2.

5. FLUX AND METABOLIC CONTROL ANALYSIS

5.0 Introduction

In Chapter 2 it was demonstrated how predictions for the metabolic engineering of a particular reaction pathway in a cell could be made on the basis of *in vitro* kinetic parameters using techniques described in Chapter 1. Chapters 3 and 4 then described the experiments and analytical techniques necessary to provide the data for the verification of these predictions. This chapter will show how such data can be used for the identification of rate-limiting enzymes by the calculation of flux control coefficients. In doing this other issues which have been raised by Chapter 4 are addressed: for example, what is the most appropriate "framework" which should be used to display and interpret fermentation data, and how can data from on-line, bulk measurements such as gas analysis be brought together with that from off-line, specific pathway data, such as intermediate metabolite concentration measurements so that a useful analysis of it can be made. A summary of the analyses performed on the data from the fermentations described in Chapter 4 is shown in Figure 5.1.

This approach is an attempt to deal with the large amounts of data of different types which are generated by a fermentation. Data can be categorised in a number of ways. Firstly, it can be on- or off-line. On-line data is typically gathered at quite short time intervals, every 3 minutes in the case of the mass spectrometer gas analysis in this study. Because of the continuous nature of the signal, noise and out-liers will be more evident. On-line data analysis tends by its nature to be rapid, so that it is reasonably certain that what is being measured is that which is being experienced by the cells. For off-line data, samples are generally taken at 30 or 60 minute intervals, with the results of the analysis being available anything from immediately in the case of O.D. to 8 hours later for dry weights. Because of possible delay between sampling and analysis it is necessary to minimize any changes which could occur in the sample from chemical biodegradation and biological metabolism. (This is normally done by reducing the temperature to at least 4°C, as described in Section 3.4.4.)

A distinction can also be made between measurements of bulk or lumped variables and those of specific variables, since each requires different treatment. In this context, bulk

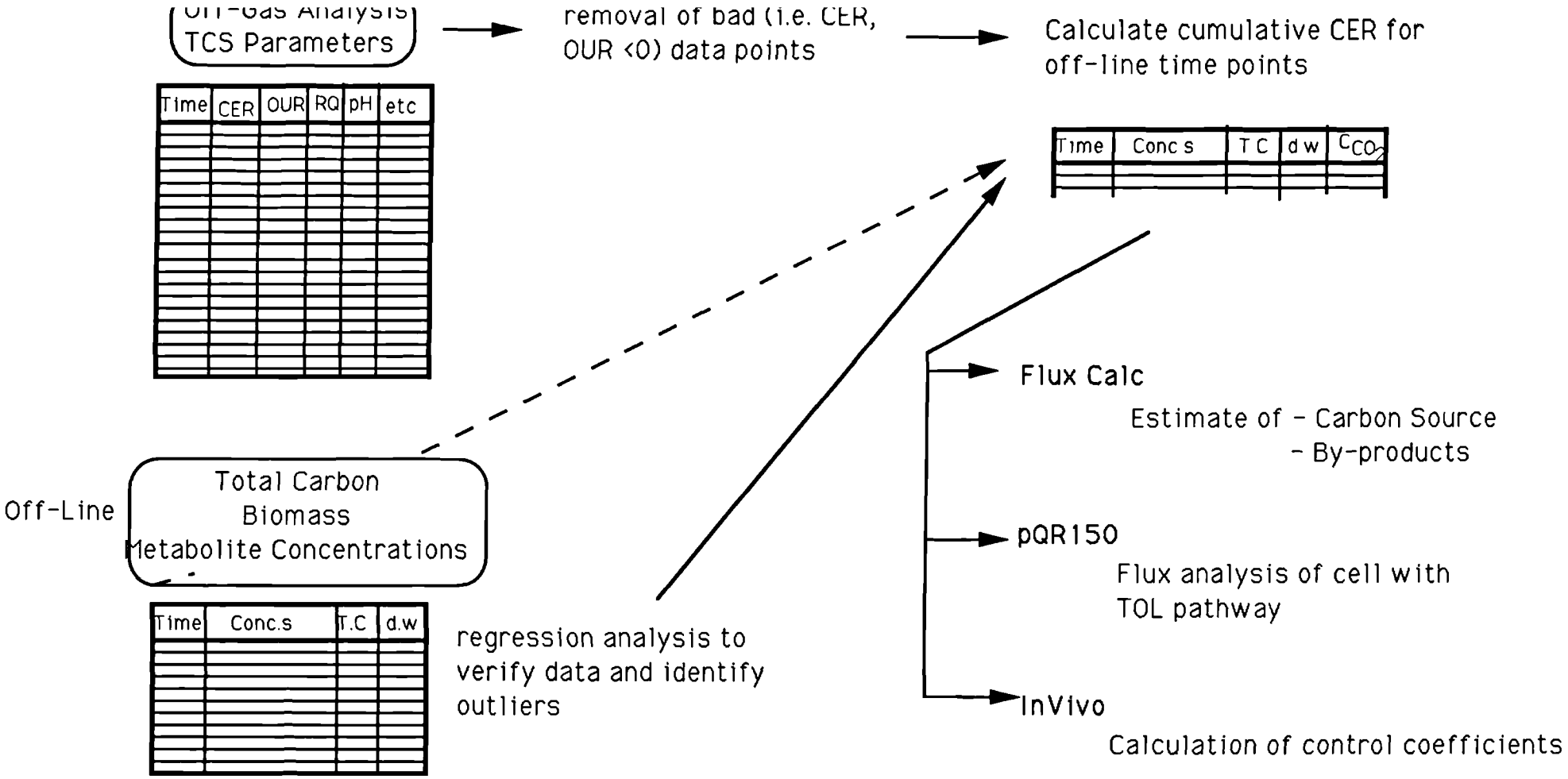


Figure 5.1: Data flow for flux and metabolic control analysis.

variables can be defined as those variables which have several input or output points in the physiological pathways or which are an agglomeration of more than one component of the system. Examples of lumped variables are total carbon, total nitrogen, CO_2 evolved and O_2 taken up and biomass. Examples of specific variable measurements are those of specific pathway substrates or intermediates, glycerol or glucose concentration, production of specific organic acids and the concentration of any products of interest.

Because of the difficulties of dealing with data which differs both in the time-frame in which it is taken and in the extent to which it is lumped, many of the models described in Section 1.3.1 restrict themselves to just one type. For example, in the fermentation models of Papoutsakis (Papoutsakis *et al*, 1986), bulk measurements of biomass and CO_2 are used, whereas in the structured model of *Bacillus subtilis* developed by Backman (Backman *et al*, 1989), concentrations of each of the pathway metabolites were required. In practical terms, however, both types of measurements are routinely carried out so that it is necessary to have a modelling framework which is flexible enough to be capable of incorporating different forms of data.

The framework outlined in this chapter for flux and metabolic control analysis is based upon specific (such as pathway intermediates) and bulk (such as CO_2 and biomass) system concentration measurements. It uses a graphical programming system called LabVIEW for data handling and flux analysis and the Mathematica based program InVivo (described in Section 2.3) for the calculation of control coefficients. Because of the modular nature of the LabView environment, it is possible to use whole existing pathway modules, to link pathway modules in different ways and to create new pathways. Data can then be fed in to calculate the fluxes through the system. The program (or Virtual Instrument, in LabView terminology) Carbon Fluxes uses lumped data (biomass, total carbon, CO_2) along with specific metabolite data (benzoate concentration) to provide estimates of carbon flux to by-products and carbon source concentration. The program pQR150 calculates the system fluxes for the case where a pathway which feeds into the central carbon pathways is introduced into a microorganism. This program illustrates the capability of the programming environment to combine new pathways with those existing.

The system fluxes calculated in this way are used for the calculation of the control

coefficients as described in the second part of the chapter.

The program Central Carbon examines the fluxes through the central carbon metabolism of E.coli HB101 growing aerobically on glucose and producing a mixture of organic acids, and how these fluxes are altered in a mutant, deficient in the enzymes for acetate production.

The second part of this chapter describes the calculation of the control coefficients for the TOL pathway from transient fermentation data. This requires a slightly different programming approach, which is described in Section 2.3 and this is explained in detail in this chapter. With improvements in graphical user interfaces and the development of software which can be run on several platforms it will be possible to combine these two approaches.

5.1 Flux Analysis

5.1.0 Introduction

The analysis of the fluxes through the metabolic pathways of an organism is a technique which can be used to process fermentation data, identify control points in a pathway and, as described below, increase fermentation productivity. By measuring the inputs and outputs from a pathway of interest, we can calculate the fluxes through the steps in the pathway for different fermentation conditions. An analysis technique of this kind can be used in several ways:

1) In the case of an already established process, flux analysis can determine what proportion of the carbon flux is being directed towards the product of interest and examine the effect upon this of changes in fermentation conditions. It may identify a need for additional analysis of by-products.

2) Where a process is at the development stage, the identification of control points and overflows to waste by-products can be used to direct genetic engineering efforts , either to overcome the regulatory mechanism or to eliminate the flow to by-product.

3) It can be used in a predictive way, for examining the effect upon the system fluxes

of the addition of new pathways to an existing system. For example, it would be possible to determine the increased demand for reducing equivalents and the subsequent changes in central carbon metabolism occurring as a result of a new pathway.
4) A flux analysis system forms a useful and intelligent framework upon which to "hang" fermentation data, that is, a wide range of fermentation parameters, from overall off-gas analysis to individual secondary metabolite concentrations can be correlated and presented using this technique, and mass balances carried out.

5.1.1 Method

Setting up the pathway:
There are three initial steps which must be followed for the flux analysis of a process:
1) Identify the important metabolic pathways or groups of metabolic pathways in the process of interest.
2) Write down all of the compounds which are or will be analysed in the process.
3) Using 1) and 2) as a basis for simplification, write down the metabolic pathway for the system. This step in itself is extremely instructive, as it may identify by-products which require analysis.
There are a range of pathway modules already constructed in the software, and these should be reviewed before programming any new pathways as novel pathways can often be constructed by altering those already present.

5.1.2 Construction of the pathway modules

In LabVIEW, because it is designed primarily as an instrumentation tool, the program modules which handle the data, describe the pathways and output the date are known as virtual instruments (VIs). A LabVIEW VI consists of a front panel, a block diagram and an icon/connector. The front panel is the user interface, the block diagram is the VI source code and the icon/connector is the calling interface. A block diagram contains input/output, computational and subVI components, which are represented by icons and interconnected by lines directing the flow of data. SubVI components call other VIs, passing data through their icon/connectors. Thus there is a hierarchy of VIs which is illustrated in Figure 5.2. The CCM (Central Carbon Metabolism) VI is the controlling

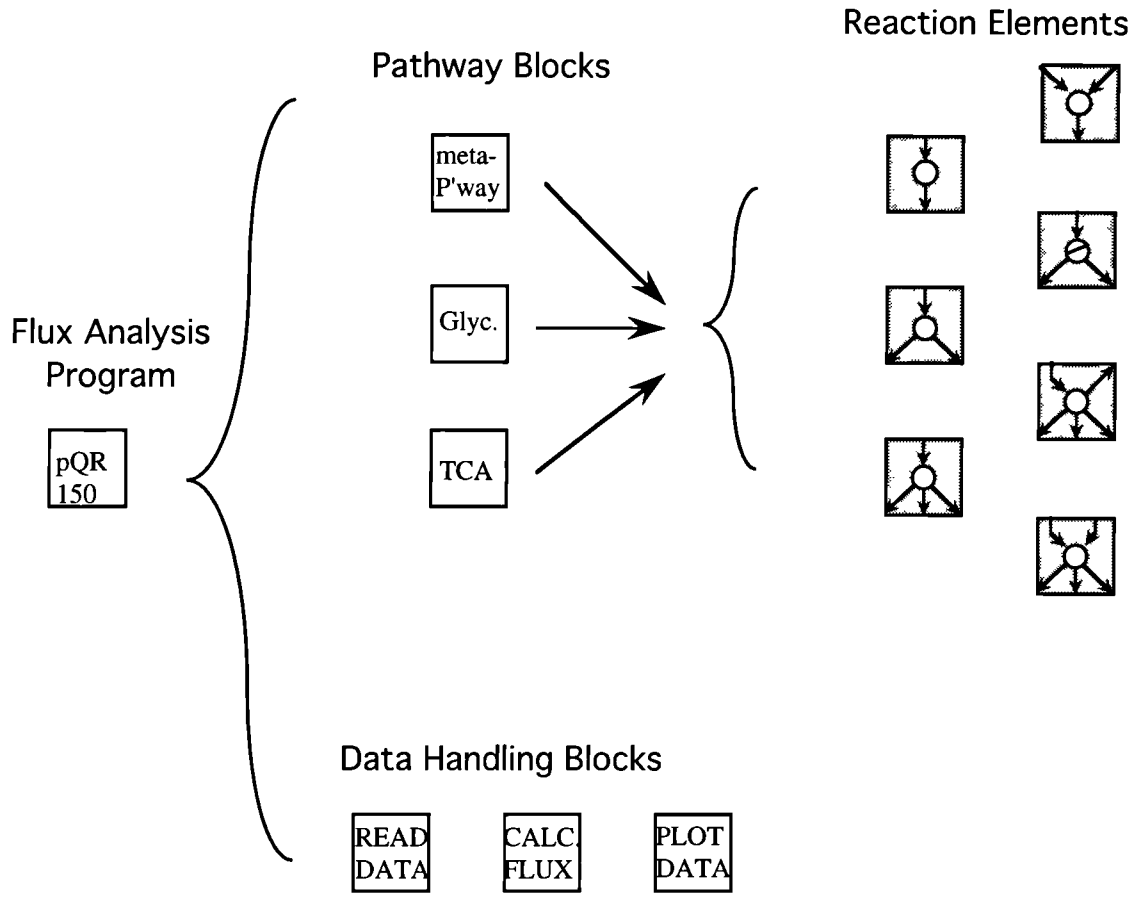


Figure 5.2: The hierarchy of the graphical elements which have been constructed in the LabView programming environment.. The reaction elements are linked to form the metabolic pathways which are separated into physiological blocks. These pathway blocks can then be linked together to reflect the interchange of metabolites between them in the cell. Associating these pathway blocks with data acquisition and data managenent blocks yields a system for the flux analysis of (in this case) E.coli JM107 pQR150.

VI in this case (more complex systems could have additional VIs such as secondary metabolite or protein production) and it contains several sub-VIs which can be grouped into three types according to their function: data reading and manipulation, pathway blocks and data output. This group of subVIs had a further set of subVIs: the data reading VI, READ DATA, uses a LabVIEW-supplied subVI which reads data from disk into an array, the FLUX VI has a subVI called calc mu for specific growth calculation, a flux calculator and a difference calculator. The pathway modules have as their subVIs a set of stoichiometric reactions, 1->1, 1->2, 2->1, 1->3, and so on. These are linked together to construct the metabolic pathways.

The data reading and manipulation VIs are READ DATA, and FLUX dp/dx.μ. Their function is to read in a file of metabolite and biomass concentration data (in text format) and to convert this to input and output fluxes. The FLUX dp/dx.μ VI contains three sub VIs: DATA F'MT for converting the concentration array to an array of concentration differences, a sub-VI for calculation of specific growth rate called calc mu, and FLUX CALC, which uses the specific growth rates and the concentration differences to calculate the fluxes. Fluxes are calculated by multiplying the throughput (mmol metabolite produced or consumed per g biomass produced) by the specific growth rate.

This flux data array is then broken up so that the appropriate sets of input and output fluxes are sent to each pathway sub-VI. It should be noted that while in physiological terms some of the fluxes are inputs (e.g. glucose) and some are outputs (biomass, acetate etc.), in data computational terms, all the data fed to the program can be called inputs. This distinction between physiological inputs and outputs and data inputs and outputs is quite important and care must be taken not to confuse the two.

5.1.3 Meta-Cleavage Pathway

The representation of the TOL meta-cleavage pathway as expressed in E.coli using the Pathway Builder program is shown in Figure 5.3 (full documentation and listings are given in Appendix D). This pathway has benzoate as its sole flux input and pyruvate,

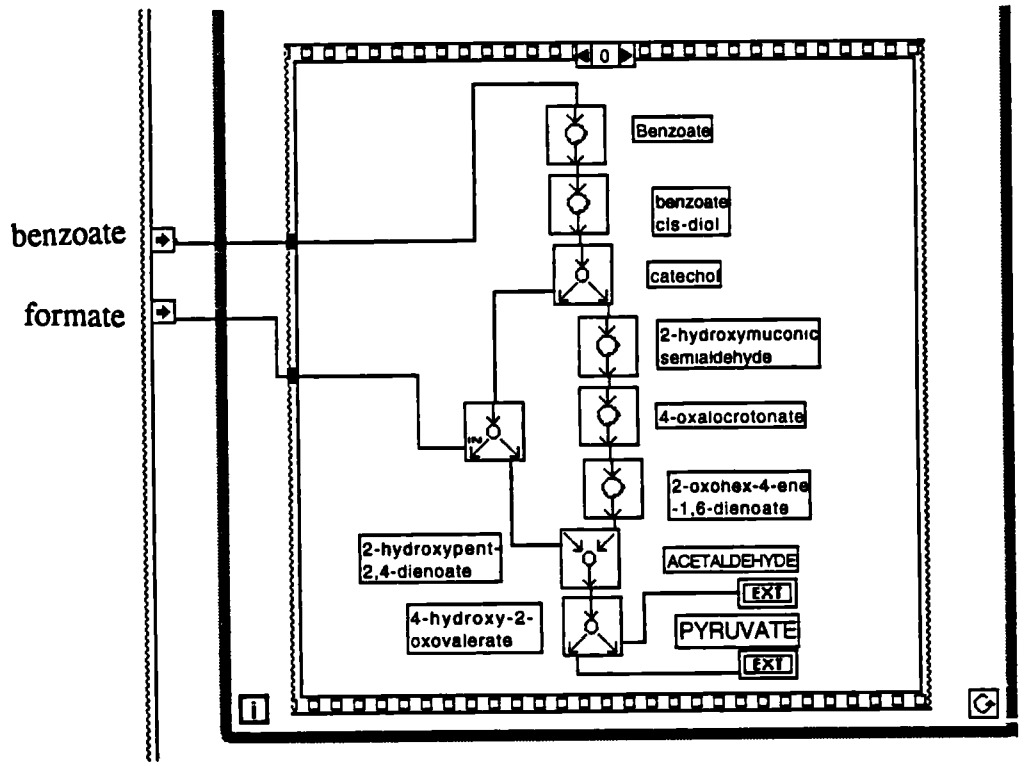


Figure 5.3: Programming code for TOL meta-cleavage pathway using graphical programming language.

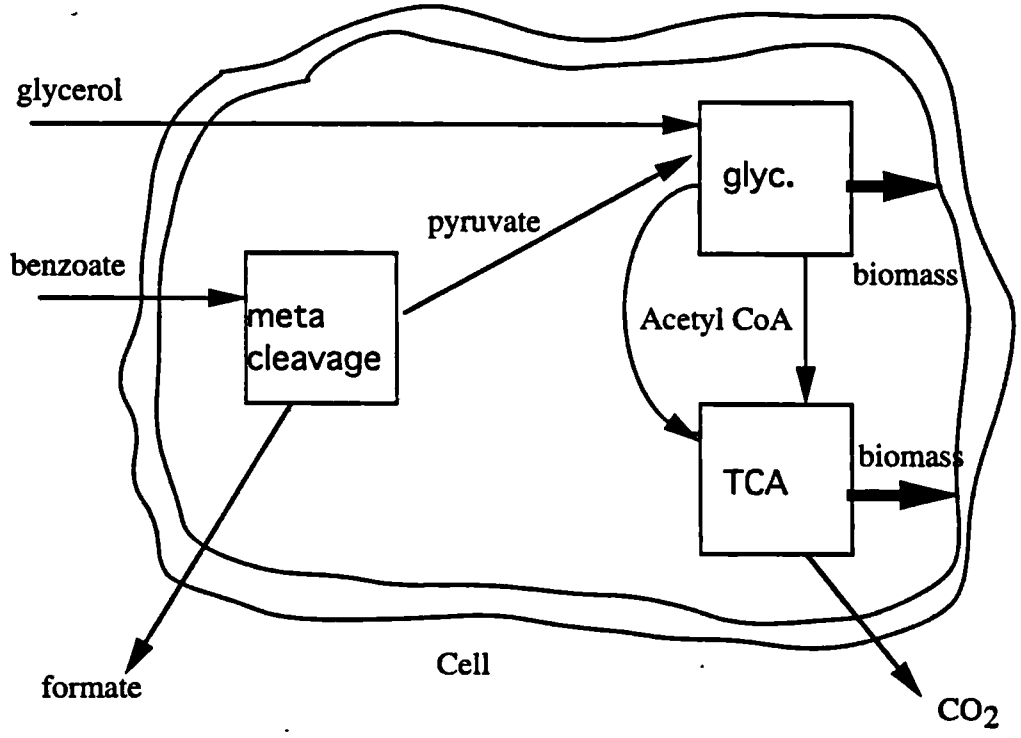


Figure 5.4: Schematic showing linking of TOL meta cleavage pathway (meta) with pathway modules for glycolysis (glyc.) and TCA (TCA).

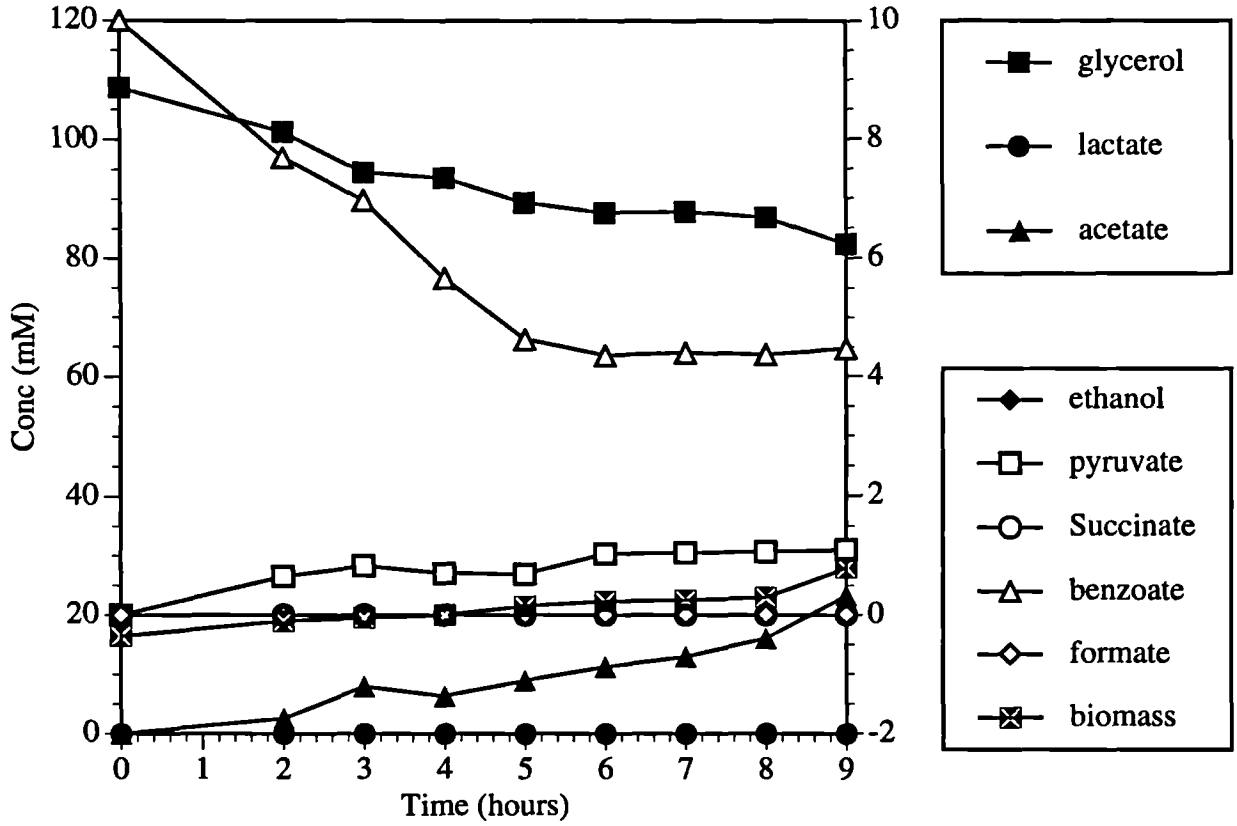


Figure 5.5: Metabolite concentrations for TOL and central carbon pathway in E.coli pQR150

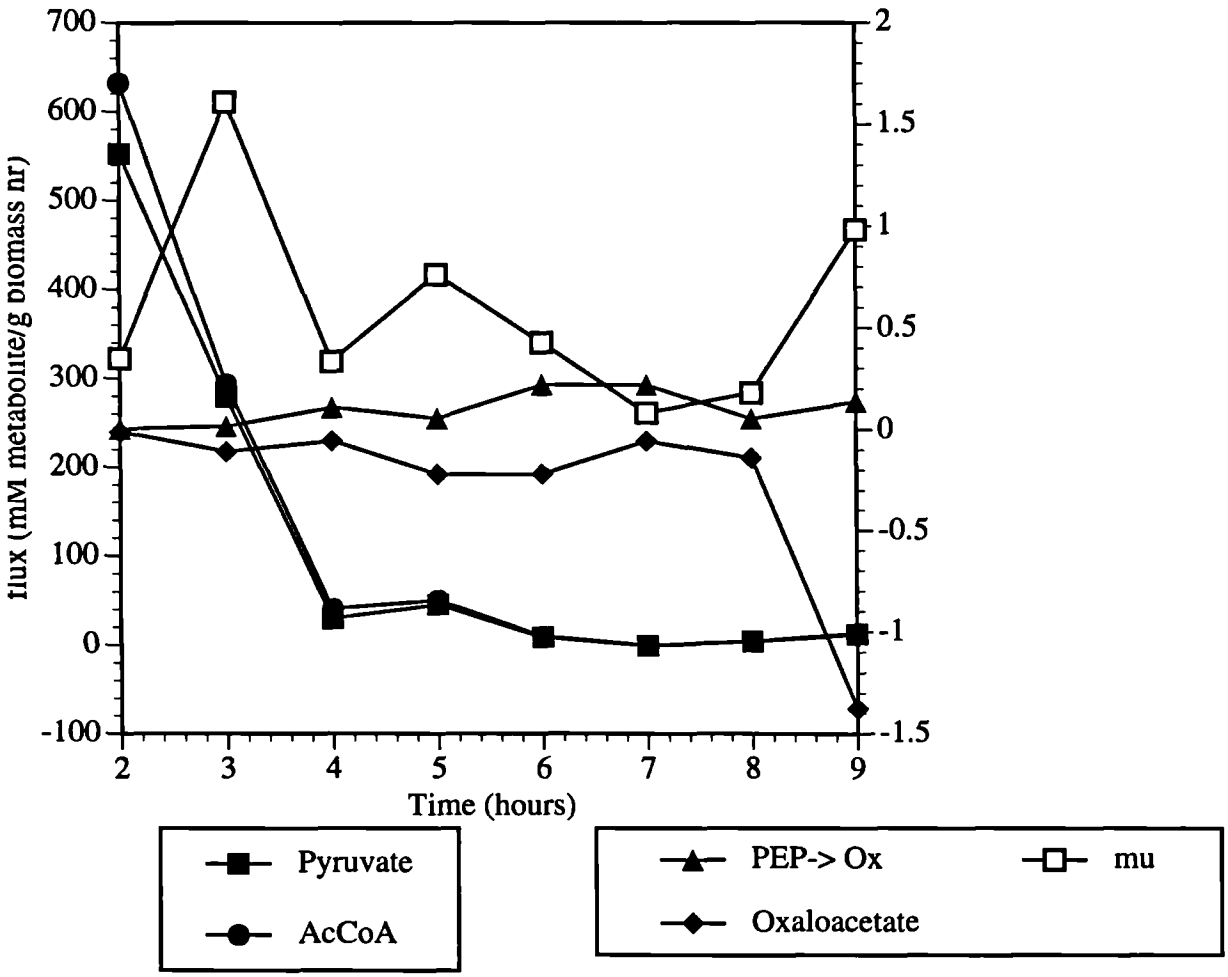


Figure 5.6: Fluxes for TOL and central carbon pathway in E.coli pQR150

acetaldehyde and formate as its flux outputs, as *E.coli* has no native mechanisms for degrading aromatics. Two parameters are required for the calculation of the fluxes in this pathway. In this case, the chosen parameters are the benzoate and formate concentrations. From these, the pyruvate and acetaldehyde fluxes are calculated and are then passed on to the central carbon metabolism VIs.

There are two central carbon metabolism VIs: one for glucogenesis and the glycolysis of glycerol to acetyl CoA, GLYC(glycerol), and the other for the TCA cycle, TCA. GLYC(glycerol) has as its inputs: glycerol, lactate, acetate, ethanol, pyruvate and acetaldehyde fluxes and differential biomass i.e. biomass$|_{t1}$-biomass$|_{t2}$. The biosynthesis stoichiometries are entered on the front panel and the defaults are those for *E.coli*. The output fluxes are those for glucose-6-phosphate, pyruvate, acetyl CoA and the flux of PEP to oxaloacetate. The last two act as inputs to the TCA VI.

The TCA VI calculates the fluxes in the TCA cycle. Its inputs are the acetyl CoA and PEP to oxaloacetate fluxes from the GLYC(glycerol) VI, the succinate flux and differential biomass. Again, the biomass stoichiometries are entered on the front panel and can be altered from it. The outputs are the oxaloacetate flux and the oxaloacetate requirement from PEP to make up for use of the TCA cycle intermediates for biomass synthesis (this last flux is passed to the GLYC VI to calculate the PEP to oxaloacetate flux). The fluxes calculated are then plotted directly onto the front panel as well as being stored as a spreadsheet file.

5.1.3.1 Results

The metabolite concentration profiles which occur during the growth of *E.coli* pQR185 on M9/IPTG/benzoate are shown in Figure 5.5. In this case the acetate levels can be calculated from the carbon flux to by-products as acetate is the only organic acid produced in detectable quantities by *E.coli* JM107 [Turner *et al*, 1994]. The benzoate concentration decreases rapidly for the first 5 hours and then remains steady due to the inactivation of pathway enzymes by the pathway intermediates (in particular that of catechol-2,3-dioxygenase by hydroxymuconic semialdehyde). The growth rate was low for this fermentation due to the inhibitory effect of benzoate and the metabolic load of plasmid expression and the final cell density was 1.4 g/l.

The metabolite fluxes for the system are shown in Figure 5.6. Because of the initial small changes in biomass as compared to glycerol uptake, the initial fluxes to G6P and pyruvate are very high but quickly drop.

The specific growth rate (μ) shows a substantial variability throughout the course of the fermentation. This is characteristic of specific growth rates calculated from dry weight measurements because of inaccuracies in the measurements and is particularly evident at the lower end of the scale as in this case. The variation in μ will cause a variation in each of the calculated fluxes as flux is defined as the product of throughput and μ. However this variation is not as evident in the other fluxes. The calculation of the specific growth rate could be improved by using the O.D. as a measure of biomass rather than the dry weight. It can be seen that when dry weight increases at 9 hours there is a corresponding increase in μ and in the flux from PEP to oxaloacetate to make up for the loss to the TCA cycle from the supply of precursors for biosynthesis. The trend in the flux from PEP to oxaloacetate follows that of the specific growth rate, but it is offset by one time point as the growth rate at one point provides the estimate for the flux to oxaloacetate needed at the next time point.

The fluxes to pyruvate and to acetaldehyde are significantly greater than the other fluxes in the system. It is a build-up of these metabolites which causes overflow to by-products, in this case, acetate. Pyruvate is also supplied to the central carbon metabolism from the TOL pathway causing high levels of intracellular pyruvate which may account for the high production of acetate seen in Figure 5.10. These relatively high pyruvate and acetyl CoA fluxes with the corresponding over-production of acetate indicate that there is a bottleneck occurring at this point.

5.1.3 Production of Organic Acids

We wish to perform a flux analysis to study the effect of blocking the pathway to acetate in E.coli K-12. This was achieved by using a strain which carried deletions in the genes *ack* and *pta*, encoding the enzymes acetate kinase and phospho-transacetylase respectively. In addition, an attempt was made to reroute the carbon flux to ethanol by

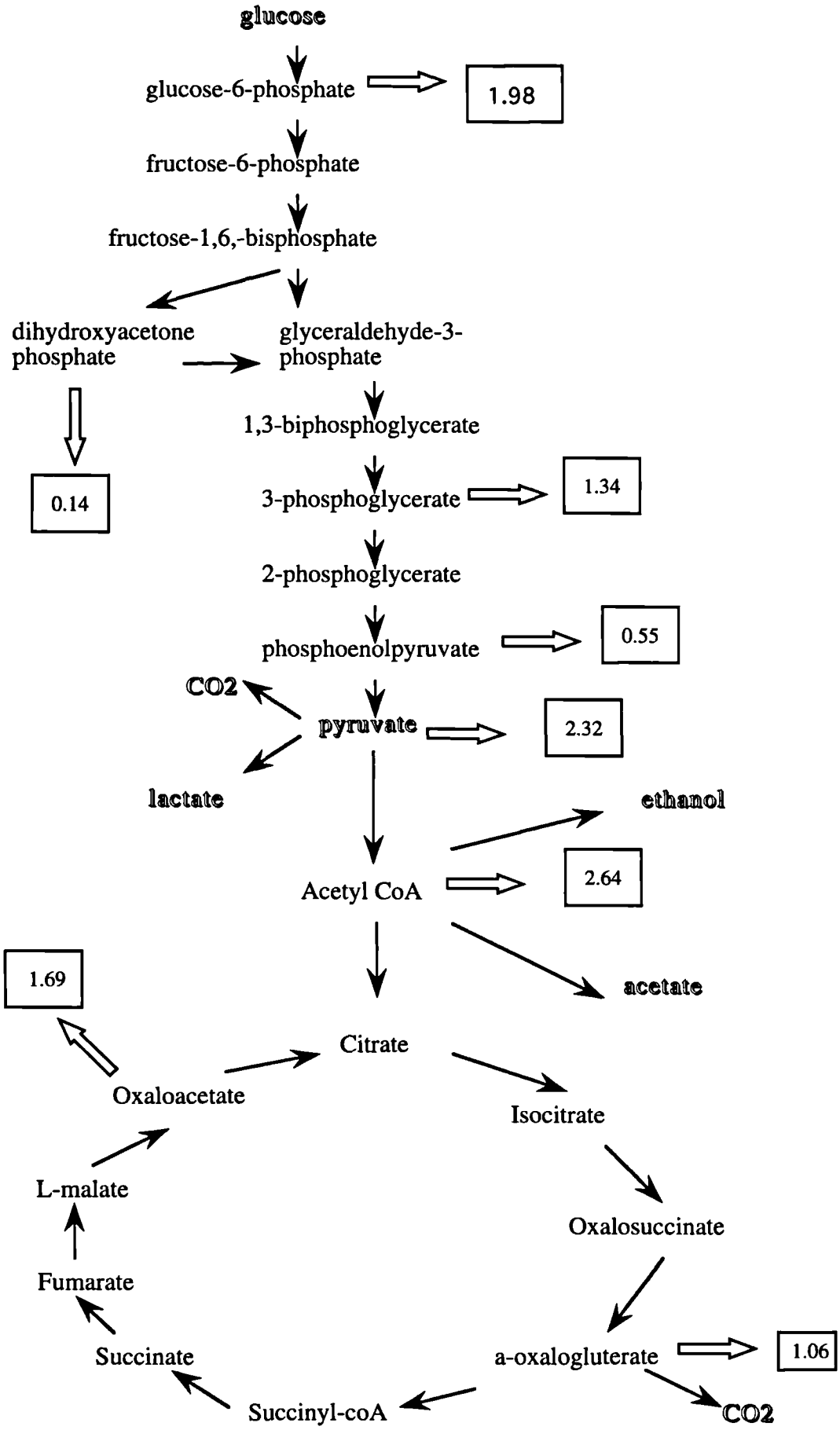


Figure 5.7: Schematic of central carbon pathway used for flux analysis, showing outputs to biomass, CO2 and byproducts.

introducing a plasmid carrying the genes for alcohol dehydrogenase. This strain produces a range of organic acids during aerobic growth on glucose, one of which is acetate.

The experiments carried out were aerobic fermentations on complex media (LB) with glucose as carbon source. Samples were taken at 30 minute intervals and were analysed for biomass, glucose, and the following organic acids: pyruvate, acetate, formate, succinate, lactate [L.Regan, M.S.Thesis, 1990, Diaz-Ricci *et al*, 1991].

From the above information it is possible to draw out a metabolic pathway. The outputs to biomass are calculated as mmol of intermediate used per g of biomass produced. These figures are calculated from the monomeric composition of *E.coli* (Dawes and Large, 1973).

The data must be entered in the following order:

time (hrs) metabolites (mmol) biomass (g/l)

The order of the metabolites will determine which numbers are used in the case loops when the appropriate data columns are cut out for the different pathway blocks (see documentation in Appendix D). The VI CCC then calculates the intermediate fluxes of the system and outputs a user-determined selection of these by means of the sub-VI PLOT FLUX.

Three cases are considered below: the first is the wild type (K-12), the second is the mutant TA3476 which is deficient in pta and ack, the third is the same mutant bearing a plasmid bearing the "pet" operon which expresses the enzymes pyruvate decarboxylase and alcohol dehydrogenase.

5.1.5.1 Results

Figure 5.8 (a) shows the metabolite concentrations for the wild type K-12 growing on glucose. Here, the primary organic acids produced are lactate, acetate and succinate. Examining the fluxes and specific growth rate μ in (b), it can be seen that the initial pyruvate flux is low. the specific growth rate reaches a maximum at 2.5 hours and afterwards decreases. At this point the pyruvate flux starts to increase, corresponding to the increasing organic acid levels in (a). The oxaloacetate flux is high initially and

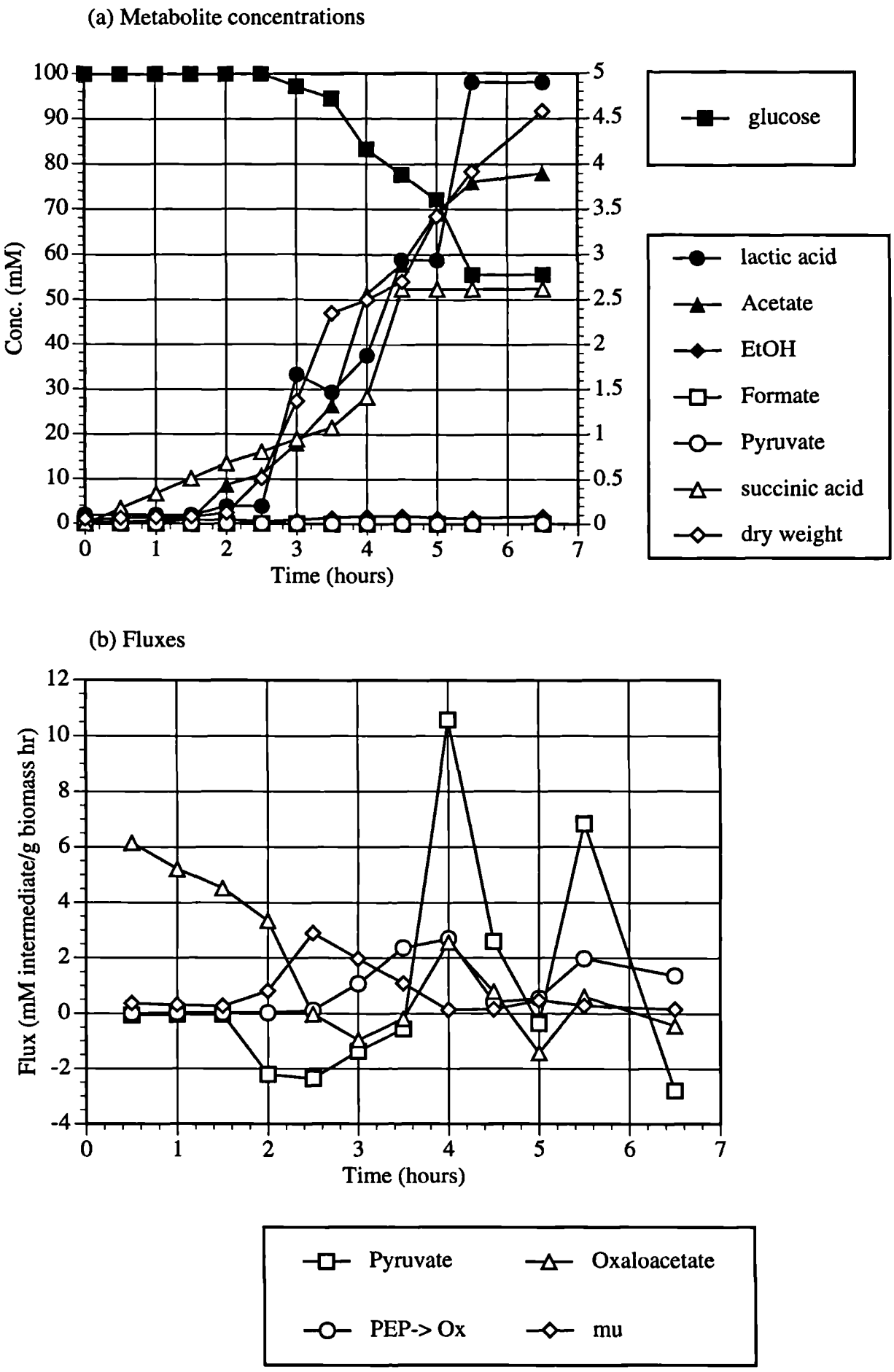


Figure 5.8: (a) Metabolite concentration profiles and (b) fluxes during growth of E.coli K-12 on glucose.

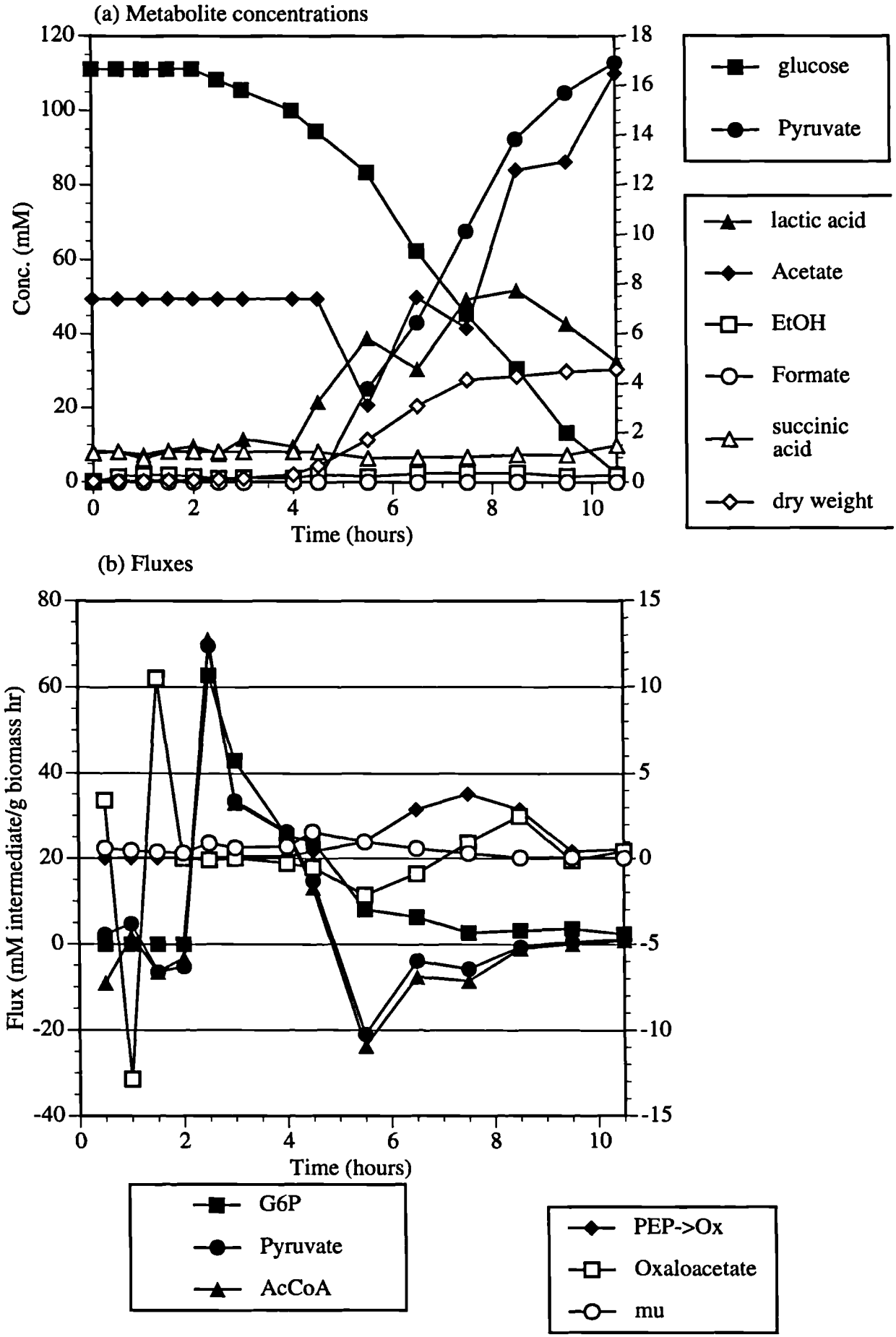


Figure 5.9: (a) Metabolite concentrations and (b) fluxes during growth of the *pta/ack* deficient mutant TA3476.

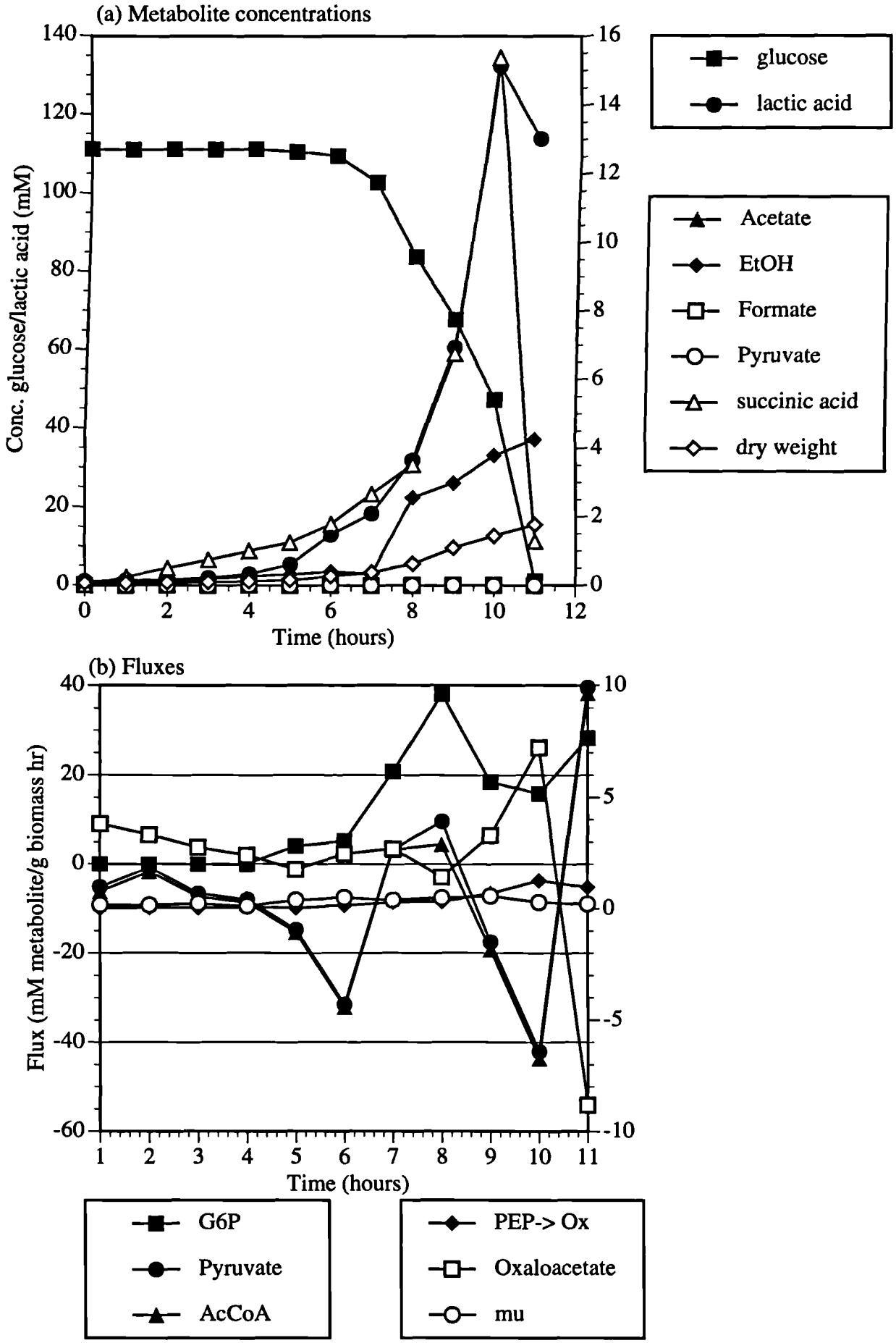


Figure 5.10: (a) Metabolite concentrations and (b) fluxes of pta/ack mutant TA3476 bearing "pet" operon

decreases as glucose utilisation increases and specific growth rises. It reaches a minimum at 3 hours and then begins to rise, after which time it follows the pyruvate flux trend. This can be explained by the fact that when glucose utilisation is low, the cell is utilising the TCA cycle. As this increases and the cell is generating enough ATP from glycolysis, the TCA cycle is disconnected and the cell produces organic acids to maintain the redox balance. As above, the flux from PEP to oxaloacetate follows the specific growth rate but offset by one data point in response to the requirement of precursors for biomass synthesis from TCA intermediates.

Comparing the performance of the *ack/pta* deficient mutant TA3476 shown in Figure 5.9, the most significant difference is in the excretion of pyruvate at 4 hours. This occurs as a result of the very large pyruvate flux seen in Figure 5.9(b) at 2.5 hours. This apparently acts as an inducer for the excretion of pyruvate and also for the expression of an alternative pathway for the production of acetate. The specific growth rate of this strain is lower than that for the wild type and reaches a maximum at 4.5 hours before decreasing. The rest of the system fluxes are comparable to those for the wild type and follow the same trend.

In Figure 5.10 it can be seen that the effect of including the *pet* operon (encoding enzymes for ethanol biosynthesis) in this mutant is to utilise these high levels of pyruvate in the production of ethanol and prevent the expression of the alternative acetate pathway. This strain produces more ethanol than either the mutant without the plasmid or the wild type. The succinic acid production is also higher. The pyruvate flux in Figure 5.10(b) is the lowest for any of the strains, showing the efficacy of this technique of rerouting the carbon flux. The oxaloacetate flux rises as succinate is excreted and then drops sharply. The specific growth rate is very low however as is the final dry weight. This indicates that just preventing the production of acetate will not necessarily lead to higher growth rates or biomass. In this case the metabolic load incurred by bearing the plasmid out-weighed the advantages of reducing the production of acetate.

The above examples illustrate how flux analysis can be used to draw conclusions about the physiology of a system from the measurement of external parameters once the reaction pathways are known. Because of the way that internal fluxes determine the external fluxes it can also be used as part of a control strategy. For example in Figure

5.9, the increased pyruvate flux causes the production of organic acids and this could be used as the basis of a control strategy.

There are several advantages to using graphical programming language in this kind of analysis. The first is the ease with which new pathways can be constructed and existing pathways altered. This is crucial when working with microorganisms whose reaction pathways may differ with a change in media or carbon source, as well as with the introduction of a plasmid bearing non-native pathways (as in the case of *E.coli* pQR150). Also it is intuitively easier to work with graphical reaction elements (as opposed to stoichiometric reaction matrices) and errors can be more easily detected. Finally it is essential that software for this kind of analysis can be interfaced to fermentation and analytical instrumentation, so that on-line analysis can be performed. This is easily done in LabView which contains application-specific libraries for data acquisition, VXI, GPIB and serial instrument control, data analysis and data storage so that a range of VI's can be constructed for the development of a complete on-line system for flux analysis.

5.2 Calculation of Control Coefficients

The previous section has dealt with the calculation of the fluxes through a metabolic system and has described how some of the regulatory characteristics of the system can be identified from a flux analysis. This method is limited, however, in that it requires qualitative judgements to be made about the importance of any individual enzymic reaction based upon observing all of the fluxes in the system. What is needed is a quantitative measurement of how influential each external concentration or enzyme level is in regulating the system such as is provided by the flux and concentration control coefficients described in Section 1.2.1. These have been calculated for the TOL pathway in a simplified and full form using kinetic data in Section 2.6 and these figures can be verified by calculation of the control coefficients using the transient concentration data obtained for the system as described in Section 4.6. The calculation of these coefficients using transient concentration data is based on the method developed by Delgado and Liao [1990, 1992] described in Section 2.4. The program InVivo described in Section 2.4 and listed and documented in Appendix D is used to calculate

the flux control coefficients from the stoichiometric matrix and the concentration data. In this section the two systems based on the TOL pathway are analysed: the first is the simplified two-step system shown in Figure 2.7, the second is the system shown in Figure 5.16.

Table 5.1: Data required for calculation of control coefficients of lumped system.

Time (hours)	Benz (mM)	Pyr (mM)	Acet (mM)	C'chol (mM)
0	10	0	0	0
2	7.7	0.65	0.65	0.01
3	6.98	0.83	0.83	0.02
4	5.66	0.71	0.71	0.02
5	4.62	0.69	0.69	0.02
6	4.36	1.04	1.04	0.01
7	4.40	1.05	1.05	0.015
8	4.37	1.07	1.07	0.01
9	4.47	1.09	1.09	0.11

5.2.1 The two-step system

This system (shown in Figure 2.4) lumps the TOL pathway into two reaction steps: one from benzoate to catechol, the other from catechol to pyruvate and acetaldehyde. Because of the ease of obtaining and analysing these compounds, lumping the system in this way allowed for a rapid control analysis of the pathway. This method allows the calculation of a control coefficient for each section of the pathway. Depending on this result it can be decided what refining of the lumping system is required: for example, if the overall control coefficient for a section of the pathway is negligible, it may not be necessary to investigate this part of the pathway further whereas a high

control coefficient for a lumped pathway indicates that additional data should be obtained to determine exactly where the controlling effect lies.

The data required for input to InVivo.m is shown in Table 5.1. It consists of concentration data for benzoate, catechol, pyruvate and acetaldehyde over a time period of 8 hours. The stoichiometric matrix for the lumped system is the same as that used in Section 2.6 and the number of non-exchangeable metabolites and the number of time points is also required as input. 'U' and 'M' are parameters required for the NAG subroutine C05NBF as described in Appendix D.

5.2.1.1 Results

The flux control coefficients for this lumped system are shown in Figure 5.11. In this figure it can be seen that the control resides primarily in the first two enzymes in the pathway. To determine how this control coefficient is partitioned among these two enzymes it is necessary to monitor the cis-diol concentration. In addition, the control coefficient for the second part of the pathway is not insignificant. This indicates that useful information can be gained by further refining the analysis of this part of the pathway by the measurement of more intermediate concentrations.

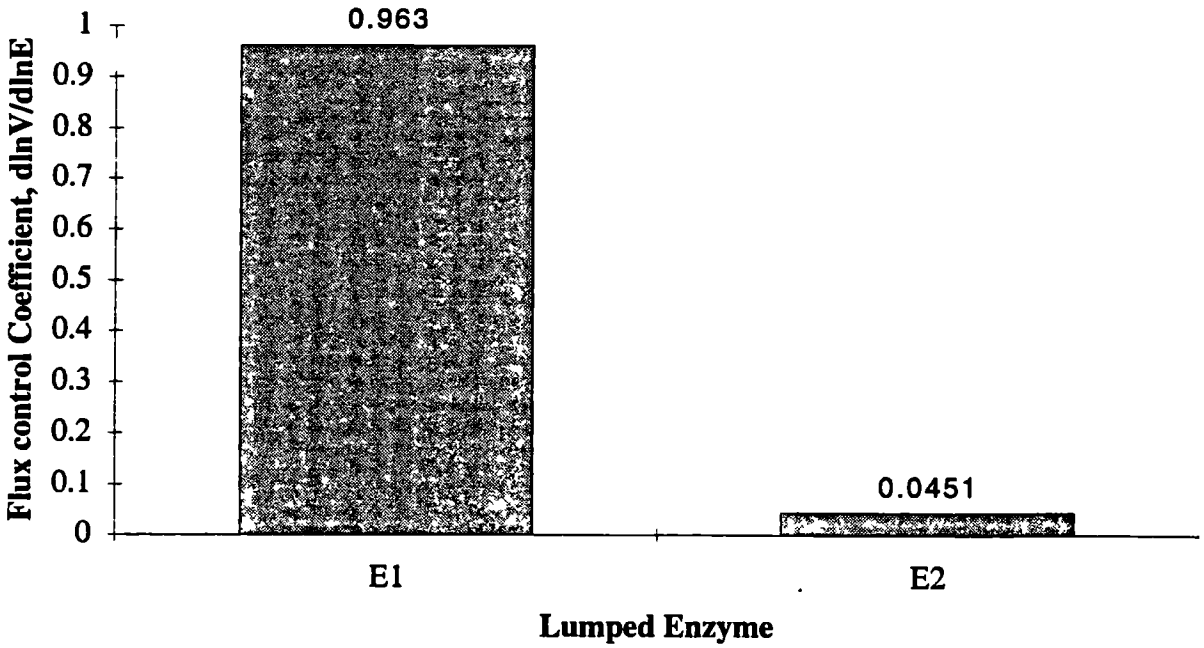


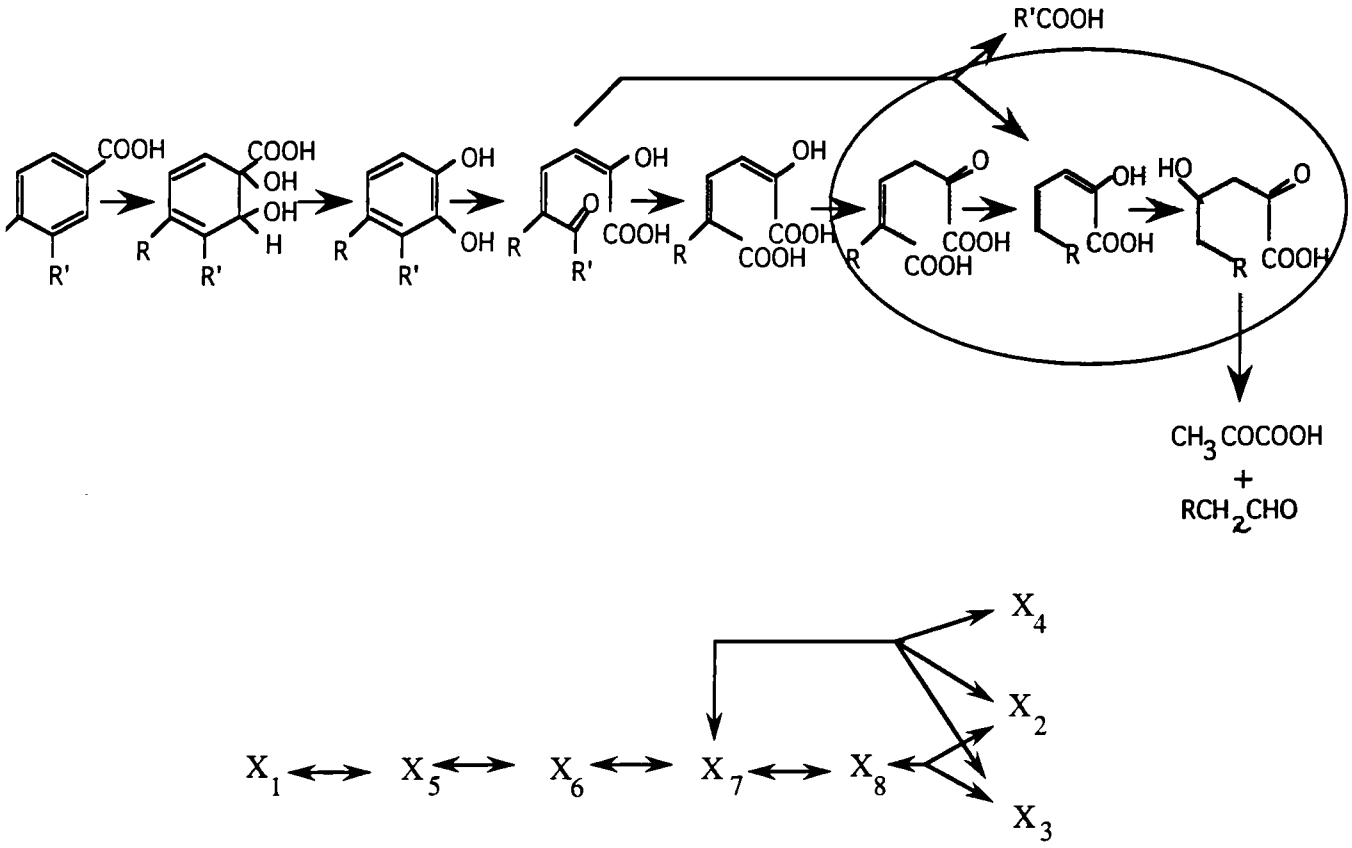
Figure 5.11: Control coefficients for the two-step system.

5.2.2 The Refined System

The results from the previous section indicate that the lumped step with the greatest control coefficient should be refined. This requires analysis of the cis-diol so that control coefficients can be assigned to the first two enzymic reactions in the TOL pathway. The refinement of the second lumped step is of less importance, as the control coefficient is smaller but it is of interest to see where the control lies. The additional metabolites in the pathway which were analysed were cis-diol, hydroxymuconic semialdehyde, oxalocrotonate and formate (see Section 3.4 and 4.5). This resulted in the more refined (but still lumped) system which is shown in Figure 5.12 with its stoichiometric matrix. The data from the intermediate concentration profiles in Figure 4.15 is used as input to InVivo.m (using the same format as shown for the two-step system in Table 5.1) along with the stoichiometric matrix in the same way as for the two-step lumped system. The results for the set of data in Figure 4.15 are shown in Figure 5.13. These are presented as a bar chart with the appropriate reactions underneath.

5.2.2.1 Results

The flux control coefficients are presented in Figure 5.13 and in Table 5.2. These indicate that the control over the pathway resides primarily in the first and second enzymes in the pathway. There is also a small control coefficient of 0.106 for 4-hydroxymuconic semialdehyde hydrolase. The control coefficient for the last 4 lumped enzymes in the pathway is small enough to indicate that it may not be necessary to make the necessary intermediate measurements to further refine the analysis of this part of the pathway. The sum of these control coefficients is 1.109 which is greater than the value of 1.0 predicted by the flux control coefficient summation theorem (see Section 1.2.1).



	r1	r2	r3	r4	r5	r6	r-1	r-2	r-3	r-4	r-5	r-6	rX1	rX2	rX3	rX4
X1	-1	0	0	0	0	0	1	0	0	0	0	0	-1	0	0	0
X2	0	0	0	1	0	1	0	0	0	-1	0	-1	0	1	0	0
X3	0	0	0	1	0	1	0	0	0	-1	0	-1	0	0	1	0
X4	0	0	0	1	0	0	0	0	0	-1	0	0	0	0	0	1
X5	1	-1	0	0	0	0	-1	1	0	0	0	0	0	0	0	0
X6	0	1	-1	0	0	0	0	-1	1	0	0	0	0	0	0	0
X7	0	0	1	-1	-1	0	0	0	-1	1	1	0	0	0	0	0
X8	0	0	0	0	0	-1	0	0	0	0	0	1	0	0	0	0

Figure 5.12: The derivation of the refined system from the TOL pathway and its stiochiometric matrix

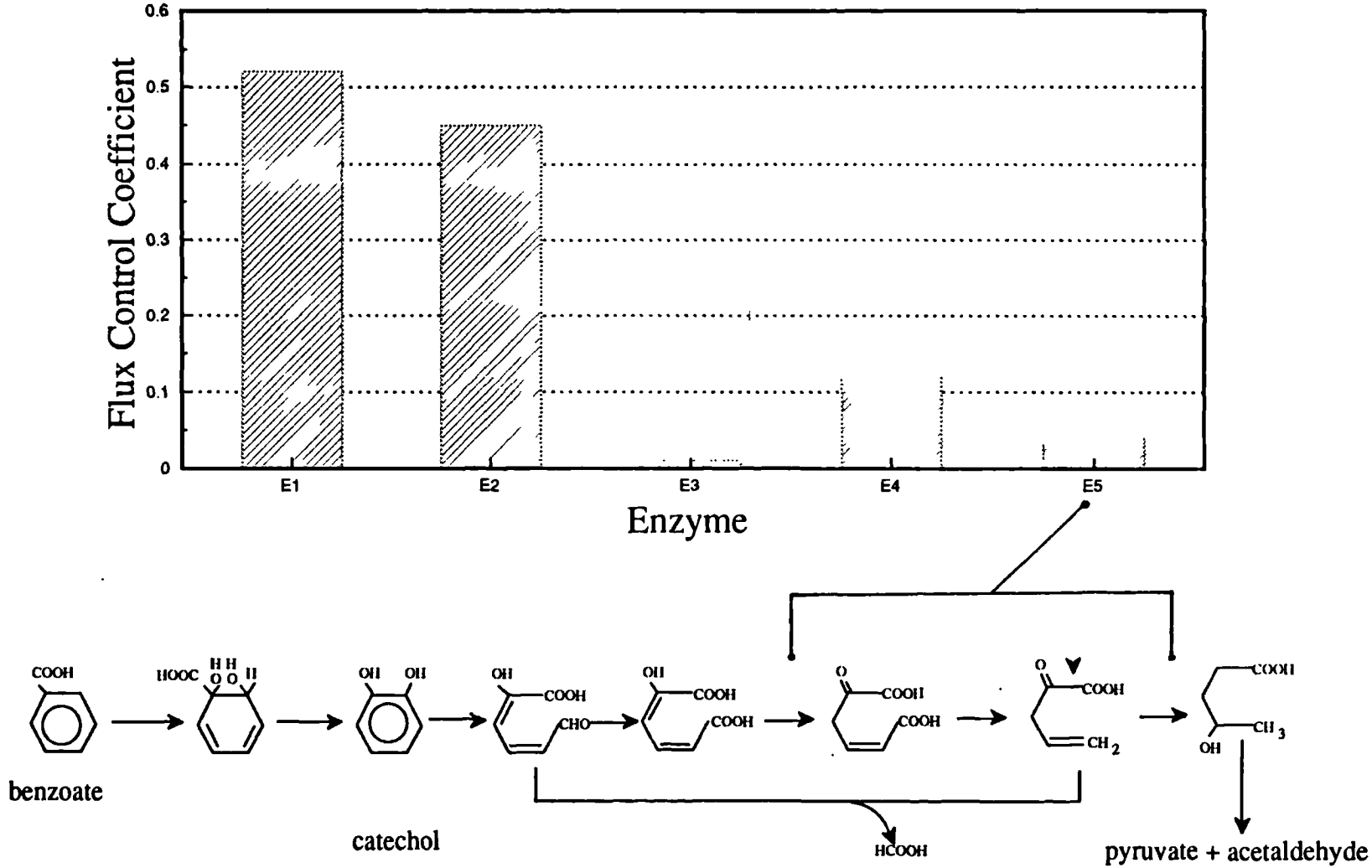


Figure 5.13: Flux control coefficients for the refined system

Table 5.2: Control coefficients for full system.

Enzyme	Control Coefficient
toluate-1,2-dioxygenase	0.508
1,2-dihydroxycyclohexa-3,5-diene-carboxylate dehydrogenase	0.463
catechol 2,3-dioxygenase	0.002
4-hydroxymuconic semialdehyde hydrolase	0.106
Lumped enzymes [4OT,4OD,OEH, HOA]	0.03

5.3 Discussion

The agreement between the lumped and the refined system is very good proving that this lumping technique can be used as a method of top-down analysis. This is useful when analysing large systems where a system may be lumped intelligently using available physiological and regulatory data. This will have the effect of high-lighting key areas in the system which are of regulatory importance. Lumping will also be necessary in cases where it may be difficult to assay certain metabolites. In this case a control coefficient can be used to determine the importance of refining the system as compared to the time and cost of developing an assay for the metabolite(s) in question. There are some differences between the system control coefficients as calculated from kinetic parameters and those from the transient concentration data. There are three significant flux control coefficients in Figure 2.9. Two of these are from the two first enzymes, TO and DCHDH, which are the rate controlling enzymes in Figure 5.13,

although the figures are higher. The third is for the enzyme 4OT which is a part of the lumped enzyme E5 in Figure 5.13 and which shows no significant control coefficient when measured using concentration data. However, the high control coefficient for OT was attributed to the functioning of the hydrolase branch of the pathway (see Figure 1.18). This is active on degradation products of benzoate in vitro but in vivo the dehydrogenase branch is used preferentially which is confirmed by the low levels of formate detected. Because of this the pathway is essentially operating as a linear pathway and the control resides in the first two enzymes.

There are several reasons why the control coefficients obtained from concentration data differ from those calculated from the kinetic data in Section 2.6.2. The first is that the kinetic parameters obtained *in vitro* are not an accurate reflection of the reaction kinetics *in vivo*. One cause of this is the deactivation of the pathway enzymes by the build-up of pathway intermediates. This will cause a change in the kinetics of the affected enzymes such as catechol-2,3-dioxygenase as the reaction proceeds and this effect has not been included in the modelling of the system in Chapter 2.

5.3 Conclusion

The modelling techniques in Chapter 2 were found to provide good qualitative predictions for the system flux control coefficients, once the difference between enzyme reaction kinetics in vitro and in vivo had been noted and accounted for. There is poor agreement in quantitative terms but this does not detract from the use of this approach to indicate areas of regulatory importance in large systems.

The results for the TOL system were as expected for such a pathway. With benzoate as substrate, it functioned in an essentially linear mode and the control resided in the first two enzymes. This indicates that the first step in engineering this pathway should be to increase the expression of TO. There is however a flux control coefficient of 0.106 for HMSD. This is significant because when the first enzyme in the pathway TO was over-expressed, there was a three-fold increase in the concentration of hydroxymuconic semialdehyde. This indicates that in the engineered pathway there may be a shift in the control features and a recalculation of the control coefficients might

implicate HMSD. This is particularly important because of the deactivation of catechol 2,3-dioxygenase by hydroxymuconic semialdehyde and leads to the consideration of non-genetic methods for flux maximization, such as the continuous removal of hydroxymuconic semialdehyde so as to keep the equilibrium on the right hand side of the forward reaction.

Secretion was not an issue in this study because of the similarity of the intra- and extra-cellular intermediate concentrations. This would not have been the case if the pathway had been studied in its native host *Pseudomonas putida*, firstly because membrane damage due to pathway intermediates would not have occurred and secondly because the function of the pathway in Pseudomonas is to supply the central carbon pathway and benzoate can be used as a sole carbon source. A similar study of the TOL pathway in Pseudomonas would require that the effect of secretion be taken into account. In addition there are several other pathways in Pseudomonas which degrade TOL pathway intermediates such as catechol (see Section 1.3.1) and the effect of these pathways would have to be included in the model. For these reasons, a control study of the TOL pathway in a host other than E.coli would not be expected to give similar results.

A vital feature of this type of analysis software is its ease of use and its compatibility with fermentation and analytical instrument. This has been addressed by using a graphical programming language to provide a front end and perform a flux analysis. With the interfacing of this to instrumentation and to the control analysis software, this provides a complete on-line system for flux and metabolic control analysis.

6.0 CONCLUSIONS AND FUTURE WORK

In this work a systematic approach has been applied to the metabolic pathway analysis of the TOL pathway. This approach consists of modelling the system using kinetic data to predict the rate-limiting enzymes, using transient fermentation data for the verification of the predictive results and then altering the pathway genetically for improved performance. To do this successfully requires a knowledge of the physiology of the system and a systematic approach to collecting and interpreting data from the system.

The software written in this study for the predictive modelling of the TOL pathway was based on a power-law approximation from biochemical systems theory and metabolic control theory and can be used for the control analysis of any metabolic system for which kinetic data is available.

A control analysis using this software indicated that the first two reaction steps in the pathway were the rate-controlling enzymes. Sensitivity analysis showed that the metabolites in the pathway most sensitive to perturbations were cis-diol and oxalocrotonate. Because the system was now approximated by a power-law formalism, it was also possible to apply linear optimization techniques to determine the set of system parameters necessary to maximize or minimize a particular system variable, such as a flux or concentration.

To verify the results obtained from a power-law approximation using kinetic data, control coefficients were calculated from transient concentration data. This required the development of analytical techniques for measurement of the concentration of the pathway intermediates. The principal analytical technique used was capillary zone electrophoresis, where the separation is by electrophoresis and the detection by spectrometry. Initially, only benzoate, catechol, pyruvate , formate and acetaldehyde could be measured so the pathway was simplified by lumping it into two reaction steps. When analysis methods were developed for oxalocrotonate and cis-diol, measurements could be made of 7 of the 9 metabolites involved in the TOL lower pathway, throughout the stationary phase of batch fermentations.

Calculation of the flux control coefficients for both the simplified and more defined pathway showed that the bulk of the control resided in the first two reaction steps. This

analysis attributed no particular sensitivity to oxalocrotonate, as had been indicated from the analysis of the kinetic data. This may have been due to the dehydrogenase branch of the pathway being inactive with the benzoate substrate. This would make the pathway strictly linear. The use of substrates which have different affinities for the dehydrogenase and hydratase branches and calculation of the resulting control coefficients would indicate if this were the case.

The use of the top-down analysis allowed rapid identification of the areas of the pathway which were not rate-controlling. This is especially important when dealing with large systems where it may be difficult to analyze every metabolite.

A systematic approach to data gathering and manipulation is also essential when dealing with large systems. This is the principal aim of the flux analysis software described in Chapter 5 which can be linked to instrumentation to provide an on-line analysis of a fermentation where the metabolic system of interest can be easily programmed by means of graphical elements.

The future work indicated by this project falls into two areas. The first is concerned with the TOL pathway. While several of the individual enzymes have been individually cloned and over-expressed, their effects on the control coefficients of the system have not yet been investigated. It is also interesting to consider the effect on the pathway of using different host organisms, in particular, the parent *Pseudomonas* strain. This would require incorporating mechanisms for secretion into the model as intra- and extra-cellular levels of the metabolites will differ.

The second area for future work is in the development of the model and software. A robust and user-friendly graphical interface has been developed using the LabView environment and this could be extended and used for on-line flux analysis.

APPENDICES

APPENDIX A: NOMENCLATURE

A	=	matrix of power-law exponents, $[a_{ij}] = g_{ij} - h_{ij}$	(m' x n)
		matrix of constraints for optimization	
b	=	vector of power-law coefficients, $b_i = \ln \beta_i/\alpha_i$	(m' x 1)
f	=	objective function	
G,H	=	matrices of power-law exponents, analogous to kinetic orders	
k	=	rate constant	
K	=	link matrix defined by $\mathbf{N} = \mathbf{K}\mathbf{N}_R$ where $\mathbf{N}_R$ is a basis set of **N** (m' x m_o)	
L	=	matrix of system metabolite sensitivities,	(m' x n)
		$[L_{ij}] = \ln X_i / \ln X_j$	
m	=	number of metabolites in system	
m'	=	number of non-exchangable metabolites	
$\mathbf{m_o}$	=	number of basis rows of **N**	
M	=	matrix of rate constant sensitivities	(m' x m')
N	=	partition of **N'** to exclude exchangable compounds	(m' x r)
N'	=	partition of **R** to exclude stoichiometry of net conversion rates	(m x r)
$\mathbf{N_R}$	=	matrix of the linearly independent basis set of **N**	(m_o x r)
n	=	number of variables affecting system	
q	=	order of reaction	
R	=	metabolic reaction matrix	(m x r+m-m')
s	=	columns of active set	
S	=	sensitivity	
$\mathbf{V_i}$	=	net flux going to increase X_i. This is an group of fluxes rather than a reaction rate.	
$\mathbf{V_{-i}}$	=	net flux going to decrease X_i	
$\boldsymbol{\nu}$	=	vector of reaction rates	(r x 1)
$\boldsymbol{\nu_A}$	=	vector of net conversion rates	(m x 1)
$\boldsymbol{\nu_{tot}}$	=	system rate vector	(r+m-m' x 1)
X_i	=	concentration of metabolite i	
y_i	=	$\ln X_i$	

α_i, β_i	=	power-law coefficients, analagous to rate constants
δ	=	increment
λ	=	Lagrange multiplier
μ	=	specific growth rate (hr^{-1})
μ_{max}	=	maximum specific growth rate (hr^{-1})
K_m	=	Michaelis-Menten parameter (mM)
k_l	=	Michaelis-Menten parameter (hr^{-1})
C^x_y	=	Control coefficient = dlnX/dlnY
A	=	Stoichiometric matrix
r	=	no. of reactions in system
B	=	matrix of atomic coefficients
q	=	no. of atomic species in system
b	=	vector of atomic species
γ	=	degree of reductance of compound
F	=	covariance matrix
J	=	flux matrix
ϵ^E_S	=	elasticity = ∂lnE/∂nS
Γ	=	steady state control matrix
τ	=	transition time for a steady state established from rest
v	=	flux
B	=	matrix of C^S_{Ej}
C	=	matrix of ϵ
dX/dt	=	the rate of change of the concentration of the pool of the metabolite X.

APPENDIX B: Some Bits of Matrix Algebra You May Have Forgotten:

It may be of use to be reminded of some common concepts in matrix algebra
A set of vectors $\{a_1,\ldots,a_n\}$in R^m is linearly independent if

$$\sum_{j=1}^{n} \alpha_j a_j = 0 \leftarrow \alpha_1 = \ldots = \alpha_n = 0$$

Otherwise a non-trivial combination of $a_1,\ldots,a_n$ is zero and $\{a_1,\ldots,a_n\}$ is said to be linearly dependent.
A subspace of R^m is a subset that is also a vector space. The set of all linear combinations of $a_1,\ldots,a_n \in R_m$ is a subset referred to as the span of $(a_1,\ldots,a_n\}$, i.e.

$$span\{a_1,\ldots,a_n\} = \{\sum_j \beta_j \alpha_j \mid \beta_1,\ldots,\beta_n \in \mathbb{R}\}$$

Two important subspaces associated with a matrix A in R^{mxn} are the range and the null space.
The range of A is defined by

$$R(A) = \{ y \in R^m \mid y = Ax \text{ for some } x \in R^n\}$$

and the null space of A by

$$N(A) = \{ x \in R^n \mid Ax = 0\}$$

If $A = [a_1,\ldots,a_n]$ then

$$R(A) = span\{a_1,\ldots a_n\}$$

The rank of a matrix A is defined by

$$rank(A) = dim[R(A)]$$

An alternative definition of rank: A matrix A is said to have rank r if it contains at least one r-rowed square submatrix with a non-zero determinant whith the determinant of any square submatrix having r+1 or more rows, possibly contained in A, is zero.
It should be noted that for any $A \epsilon R^{mxn}$, dim[N(A)] + rank(A) = n, and that rank(A) = rank(A^T) so that the rank of a matrix equals the maximal number of independent rows or columns.

APPENDIX C

C.1 Number of reactions required to determine system: (Niranjan and San 1989)

If we divide the m compounds into s external variables and p internal variables, where the external variables are those which are exchanged by the system and the internal variables are those which are internal to the system, we can partition **A** into two subsystems

$$[A_{ext}^T] = n_{ext} \tag{C1}$$

$$[A_{int}^T] = n_{int} \tag{C2}$$

Using a quasi-steady state assumption on the internals (that is, assuming that they are non-accumulating) we have

$$[A_{int}^T]x = 0 \tag{C3}$$

Subjecting Eqn. C1 to the q elemental balances we can reduce it to

$$[\tilde{A}_{ext}^T]x = \tilde{n}_{ext} \tag{C4}$$

where $\tilde{\mathbf{A}}^T_{ext}$ is an (s-q) x r matrix. Dividing the right hand side of this equation into two

parts we can rewrite it as

$$\begin{bmatrix} \tilde{A}_{ext}^{T} & 0 \\ & -I_l \end{bmatrix} \begin{bmatrix} x \\ \tilde{l} \\ n_{ext} \end{bmatrix} = \begin{bmatrix} \tilde{s-q-l} \\ n_{ext} \\ 0 \end{bmatrix} \tag{C5}$$

Similarly Eqn. C3 becomes

$$\begin{bmatrix} A_{int}^{T} & 0 \end{bmatrix} \begin{bmatrix} x \\ \tilde{l} \\ n_{ext} \end{bmatrix} = [0] \tag{C6}$$

Here, n_{ext}^{s-q-l} represents the fluxes of the externals that are externally measured.
If the r reactions are independent (or the system has been reduced to ensure this) then the coefficient matrix in Eqn. C6 has linearly independent columns and its rank is r + l. Because of this it follows that

$$r + l = s - q + p \tag{C7}$$

is a necessary and sufficient condition to solve the system uniquely for any values of the s - q - l experimentally determined external fluxes [Tsai and Lee 1988].
When $r = r_{max}$, $l = 0$ and the only relation among the s external metabolites is that derived from the elemental balances.

C.2 Flux-Oriented Theory - a Treatment of Substrate Cycling: (Crabtree and Newsholme, 1985)

Using the product rule for series reactions

$$\begin{aligned} s_{i(X)}^{E_2} &= r_{(X)}^{F}(1+C/J) &&= a \\ s_{i(S)}^{E_2} &= r_{(S)}^{F}(1+C/J) &&= b \\ s_{i(P)}^{E_2} &= f_{(P)}^{F}(-C/J) &&= c \end{aligned} \tag{C8}$$

Using the power-law formalism we can express the reaction rates as

$$\begin{aligned} v_1 &= k[S]^a[P]^b[X]^c \\ v_2 &= k[S]^d \\ v_3 &= k[P]^e \end{aligned} \tag{C9}$$

In steady state $v_1=v_2=v_3=J$ so that substituting into Eqn. 8 the expressions for [S] and [P] we have

$$J = k[X]^{s^J_{(X)}} \tag{C10}$$

where

$$s^J_{(X)} = \frac{s^{E_1}_{i(S)} s^{E_2}_{i(P)} r^F_{(X)} (1+C/J)}{\left(s^{E_1}_{i(S)} s^{E_3}_{i(P)} - s^{E_3}_{i(P)} r^F_{(S)} (1+C/J) + s^{E_1}_{i(S)} r^C_{(P)} C/J \right)} \tag{C11}$$

We can use this expression to derive the conditions under which cycling will increase the sensitivity of the flux to the regulator X, as follows:

For zero cycling (C equals zero)

$$\begin{aligned} s_o &= \frac{s^{E_1}_{i(S)} s^{E_3}_{i(P)} r^F_{(X)}}{s^{E_1}_{i(S)} s^{E_3}_{i(P)} - s^{E_3}_{i(P)} r^F_{(S)}} \\ &= \frac{s^{E_1}_{i(S)} r^F_{(X)}}{s^{E_1}_{i(S)} - r^F_{(S)}} \end{aligned} \tag{C12}$$

For infinite cycling (C equals ∞)

$$s_{\infty} = \frac{s_{i(S)}^{E_1} s_{i(P)}^{E_3} f_{(X)}^{F}}{s_{i(S)}^{E_1} r_{(P)}^{C} - s_{i(P)}^{E_3} r_{(S)}^{F}} \tag{C13}$$

$$p = \frac{s_o}{s_{\infty}} = \frac{(s_{i(S)}^{E_1} - r_{(S)}^{F}) s_{i(P)}^{E_3}}{s_{i(S)}^{E_1} r_{(P)}^{C} - s_{i(P)}^{E_3} r_{(S)}^{F}} \tag{C14}$$

All the terms in this expression are positive except $s^{E1}_{i(S)}$ (since this represents an inhibition) and, replacing this by its' absolute value $s^{E1}_{i(S)}$, we obtain

$$p = \frac{(s_{i(S)}^{-E_1} + r_{(S)}^{F}) s_{i(P)}^{E_3}}{s_{i(S)}^{-E_1} r_{(P)}^{C} + s_{i(P)}^{E_3} r_{(S)}^{F}} \tag{C15}$$

For cycling to increase sensitivity we require $p > 1$

$$s_{i(P)}^{E_3} [s_{i(S)}^{-E_1} + r_{(S)}^{F}] - (s_{i(S)}^{-E_1} r_{(P)}^{C} + s_{i(P)}^{E_3} r_{(S)}^{F}) > 0 \tag{C16}$$

$$\rightarrow \quad s_{i(S)}^{-E_1} [s_{i(P)}^{E_3} - r_{(P)}^{C}] > 0 \tag{C17}$$

$s^{Ei}_{i(S)} > 0$ so that our condition for cycling to increase the sensitivity of the flux to X is

$$s_{i(P)}^{E_3} > r_{(P)}^{C} \tag{C18}$$

That is, if the sensitivity of the reaction catalyzed by E_3 to P is greater than the sensitivity of the reverse cycle to P.

C.3 Model Differential System:

(Reder, 1988)

$$\zeta(\mu): \frac{dx}{dt} = N v(x;\mu) \tag{C19}$$

where μ = external parameter vector

$\mathbf{x}$ = concentration vector

$\mathbf{v}$ = reaction rate vector

$\mathbf{N}$ = stoichiometric matrix

At a steady state designated σ°

$$\mathbf{N}\ \mathbf{v}(\sigma^\circ;\mu^\circ) = \mathbf{0} \tag{C20}$$

and we assume that ($\mathbf{N_R}\ D_x\mathbf{v}\ \mathbf{L}$) is invertible.

Defining the steady state flux function J

$$\mathbf{J}\ (\chi;\mu) = \mathbf{v}\ (\sigma(\chi;\mu):\mu) \tag{C21}$$

so that **J** associates with every couple $(\chi;\mu)$ the rate vector corresponding to the steady state $\sigma(\chi;\mu)$.

We can define the rate vector **V** in a similar way

$$\mathbf{V}\ (\chi;\mu) = \{\ \mathbf{v}(\chi;\mu)\ ;\ \chi \in \mathcal{L}(\chi)\ \} \tag{C22}$$

There exists a simple steady state control matrix Γ such that

$$D_\mu\sigma = \Gamma\ D_\mu\mathbf{v} \tag{C23}$$

$$D_\chi\sigma = \mathbf{I}_m + \Gamma\ D_\chi\mathbf{v} \tag{C24}$$

Similarly there exists a simple steady state flux control matrix **C** such that

$$D\mu\mathbf{J} = \mathbf{C}\, D\mu\mathbf{v} \tag{C25}$$

$$D\chi\mathbf{J} = \mathbf{C}\, D\chi\mathbf{v} \tag{C26}$$

Γ and **C** can be obtained from the expressions

$$\mathbf{\Gamma} = -\mathbf{L}\,(\mathbf{N}_R\, D_x\mathbf{v}\, \mathbf{L})^{-1}\mathbf{N}_R \tag{C27}$$

$$\mathbf{C} = \mathbf{I}_r - D_x\mathbf{v}\, \mathbf{L}\,(\mathbf{N}_R\, D_x\mathbf{v}\, \mathbf{L})^{-1}\mathbf{N}_R \tag{C28}$$

$$\Gamma_{ij} = [\frac{\partial \sigma_i}{\partial \lambda_j}(\sigma^o;\lambda^o)][\frac{\partial v_j}{\partial \lambda_j}(\sigma^o;\lambda^o)]^{-1} \tag{C29}$$

$$C_{ij} = [\frac{\partial J_i}{\partial \lambda_j}(\sigma^o;\lambda^o)][\frac{\partial v_j}{\partial \lambda_j}(\sigma^o;\lambda^o)]^{-1} \tag{C30}$$

where λ_j is a parameter that acts on v_j only at point $(\sigma^o;\lambda^o)$

C.4 Proof for Linear System

(Westerhoff and Chen, 1984)

$$\underset{}{\overset{E_1 \quad E_2 \quad E_3}{A \rightleftarrows S_1 \rightleftarrows S_2 \rightleftarrows P}}$$

the above theorems can be written as

$$C_{E_1}^{J} + C_{E_2}^{J} + C_{E_3}^{J} = 0 \tag{87}$$

$$C_{E_1}^{J}\epsilon_{S_1}^{E_1} + C_{E_2}^{J}\epsilon_{S_1}^{E_2} + C_{E_3}^{J}\epsilon_{S_1}^{E_3} = 0 \tag{88}$$

$$C_{E_1}^{J}\epsilon_{S_2}^{E_1} + C_{E_2}^{J}\epsilon_{S_2}^{E_2} + C_{E_3}^{J}\epsilon_{S_2}^{E_3} = 0 \tag{89}$$

$$C_{E_1}^{S_1} + C_{E_2}^{S_1} + C_{E_3}^{S_1} = 0 \tag{90}$$

$$C_{E_1}^{S_2} + C_{E_2}^{S_2} + C_{E_3}^{S_2} = 0 \tag{91}$$

$$C_{E_1}^{S_1}\epsilon_{S_1}^{E_1} + C_{E_2}^{S_1}\epsilon_{S_1}^{E_2} + C_{E_3}^{S_1}\epsilon_{S_1}^{E_3} = -1 \tag{92}$$

$$C_{E_1}^{S_2}\epsilon_{S_2}^{E_1} + C_{E_2}^{S_2}\epsilon_{S_2}^{E_2} + C_{E_3}^{S_2}\epsilon_{S_2}^{E_3} = -1 \tag{93}$$

$$C_{E_1}^{S_1}\epsilon_{S_2}^{E_1} + C_{E_2}^{S_1}\epsilon_{S_2}^{E_2} + C_{E_3}^{S_1}\epsilon_{S_2}^{E_3} = 0 \tag{94}$$

C.5 Theorems of Metabolic Control

These give a set of relationships between the control coefficients and the elasticities which can be developed into what are frequently called the theorems of metabolic

control. We will state these theorems below. Their physical significance will be discussed in Section 1.2.3.

1.Flux Control Summation Theorem:

$$\sum_{i=1}^{n} C_{E_i}^{J} = 1 \tag{60}$$

(from the product of row 1 of B and column 1 of A)

2.Flux Control Connectivity theorem:

$$\sum_{i=1}^{n} C_{E_i}^{J} \cdot \epsilon_{S_k}^{E_i} = 0 \tag{61}$$

(from the product of row 1 of B and columns 2 to n of A)

3.Concentration-Control Coefficient Summation Theorem:

$$\sum_{i=1}^{n} C_{E_i}^{S_j} = 0 \tag{62}$$

(from the product of rows 2 to n of B and column 1 of A)

4.Concentration-Control Coefficient Connectivity Theorem:

$$\sum_{i=1}^{n} C_{E_i}^{S_j} \cdot \epsilon_{S_k}^{E_i} = -\delta_{jk} \tag{63}$$

(from the product of rows 2 to n of B and columns 2 to n of A)

Derivation of Control Theorems from Biochemical Systems Theory:

We can obtain a set of results which is similar to the above but more general from biochemical systems theory. Referring back to Eqn. 11 but considering a system with no external variables i.e. n=0 we can write

$$y = Mb \tag{64}$$

$$\rightarrow \quad M = A^{-1} \tag{65}$$

and since

$$MA = I \tag{66}$$

$$\sum_{j=1}^{n} M_{ij} a_{jk} = \delta_{ik} \quad = \begin{cases} 1 & i = k \\ 0 & i \neq k \end{cases} \tag{67}$$

Summation and use of Eqn. 28 yields

$$\sum_{j=1}^{n} [\, S(X_k,\alpha_j) + S(X_k,\beta_j)] = 0 \tag{68}$$

This is a general constraint but can be expressed in terms of the control coefficients defined in Eqn. 46 if we consider an unbranched system with homogenous rate equations as before. In this way we obtain the concentration-control coefficient summation theorem

$$\sum_{j=1}^{2n} C_{E_j}^{X_k} = 0 \tag{69}$$

where

$$C_{E_j}^{X_k} = S(X_k,\alpha_j) \quad or \quad S(X_k,\beta_j) \tag{70}$$

Substitution of Eqn. 28 into Eqn. 70 yields

$$\sum_{j=1}^{n} S(X_i,\beta_j)a_{jk} = \delta_{ik} \tag{71}$$

$$\sum_{j=1}^{n} [\, S(X_i,\beta_j)g_{jk} - S(X_i,\beta_j)h_{jk}] = \delta_{ik} \tag{72}$$

$$\sum_{j=1}^{n} [\, S(X_i,\alpha_j)g_{jk} + S(X_i,\beta_j)h_{jk}] = -\delta_{ik} \tag{73}$$

Again for the special case defined above this reduces to the concentration-control coefficient connectivity theorem.

$$\sum_{j=1}^{2n} C_{E_j}^{X_i} \epsilon_{X_k}^{j} = -\delta_{ik} \tag{74}$$

In order to derive the flux control summation and connectivity theorems we must define a sensitivity with respect to flux similar to that in Eqn. 47 but using the parameters of biochemical systems theory. We can write a flux V_k through a pool X_k as

$$V_k = \prod_{j=1}^{n} X_j^{g_{jk}} \tag{75}$$

so that the sensitivity can be written as

$$S(V_k,\alpha_i) = \frac{\partial Ln\, V_k}{\partial Ln\, \alpha_i} = \delta_{ik} + \sum_{j=1}^{n} g_{kj} S(X_j,\alpha_i) \tag{76}$$

$$S(V_k,\beta_i) = \frac{\partial Ln\, V_k}{\partial Ln\, \beta_i} = \sum_{j=1}^{n} g_{kj} S(X_j,\beta_i) \tag{77}$$

Summing these terms,

$$\sum_{i=1}^{n} [\, S(V_k,\alpha_i) \,+\, S(V_k,\beta_i) \,] \;=\; 1+\sum_{i=1}^{n}\sum_{j=1}^{n} g_{kj}[\, S(X_j,\alpha_i) \,+\, S(X_j,\beta_i) \,] \;=\; 1 \tag{78}$$

so that, when as before we assume the special case of an unbranched system whose rate laws are homogenous with respect to the system enzymes, we find that this expression simplifies to the flux control summation theorem.

$$\sum_{i=1}^{2n} C_{E_i}^{J_k} \;=\; 1 \tag{79}$$

We can obtain a general flux control connectivity theorem as follows,

$$\sum_{i=1}^{n} [\, S(V_k,\alpha_i) g_{im} \,+\, S(V_k,\beta_i) h_{im} \,] \tag{80}$$

$$\begin{aligned}
&= \sum_{i=1}^{n} \{[\delta_{ik} + \sum_{j=1}^{n} g_{kj} S(X_j,\alpha_i)] g_{im} + [\sum_{j=1}^{n} g_{kj} S(X_j,\beta_i)] h_{im}\} \\
&= g_{km} - \sum_{i=1}^{n}\sum_{j=1}^{n} g_{kj} S(X_j,\beta_i) a_{im} \\
&= g_{km} - \sum_{j=1}^{n} g_{kj} [\sum_{i=1}^{n} S(X_j,\beta_i) a_{im}] \\
&= g_{km} - \sum_{j=1}^{n} g_{kj} [\delta_{jm}] \;=\; 0
\end{aligned} \tag{51}$$

Again, simplifying to the above special case, we obtain,

$$\sum_{i=1}^{2n} C_{E_i}^{J_k} \epsilon_{X_m}^{i} \;=\; 0 \tag{82}$$

C.6 Control coefficients for different reaction mechanisms:

Irreversible Inhibition

$$C_{E_i}^{J} = -\frac{I_{max}}{J}\cdot\left(\frac{\partial J}{\partial [I]}\right)_{I=0}$$

Competitive Inhibition

reversible reaction

$$C_{E_i}^{J} = \frac{1}{J}\left(\frac{\partial J}{\partial [I]}\right)_{I=0}\cdot -K_i\left(1 + \frac{[S]}{K_s} + \frac{[P]}{K_p}\right)$$

irreversible reaction

$$C_{E_i}^{J} = \frac{1}{J}\left(\frac{\partial J}{\partial [I]}\right)_{I=0}\cdot -K_i\left(1 + \frac{[S]}{K_s}\right)$$

Uncompetitive Inhibition

irreversible reaction

$$C_{E_i}^{J} = \frac{1}{J}\left(\frac{\partial J}{\partial [I]}\right)_{I=0}\cdot \frac{-K_s K_i}{[S]}\left(1 + \frac{[S]}{K_s}\right)$$

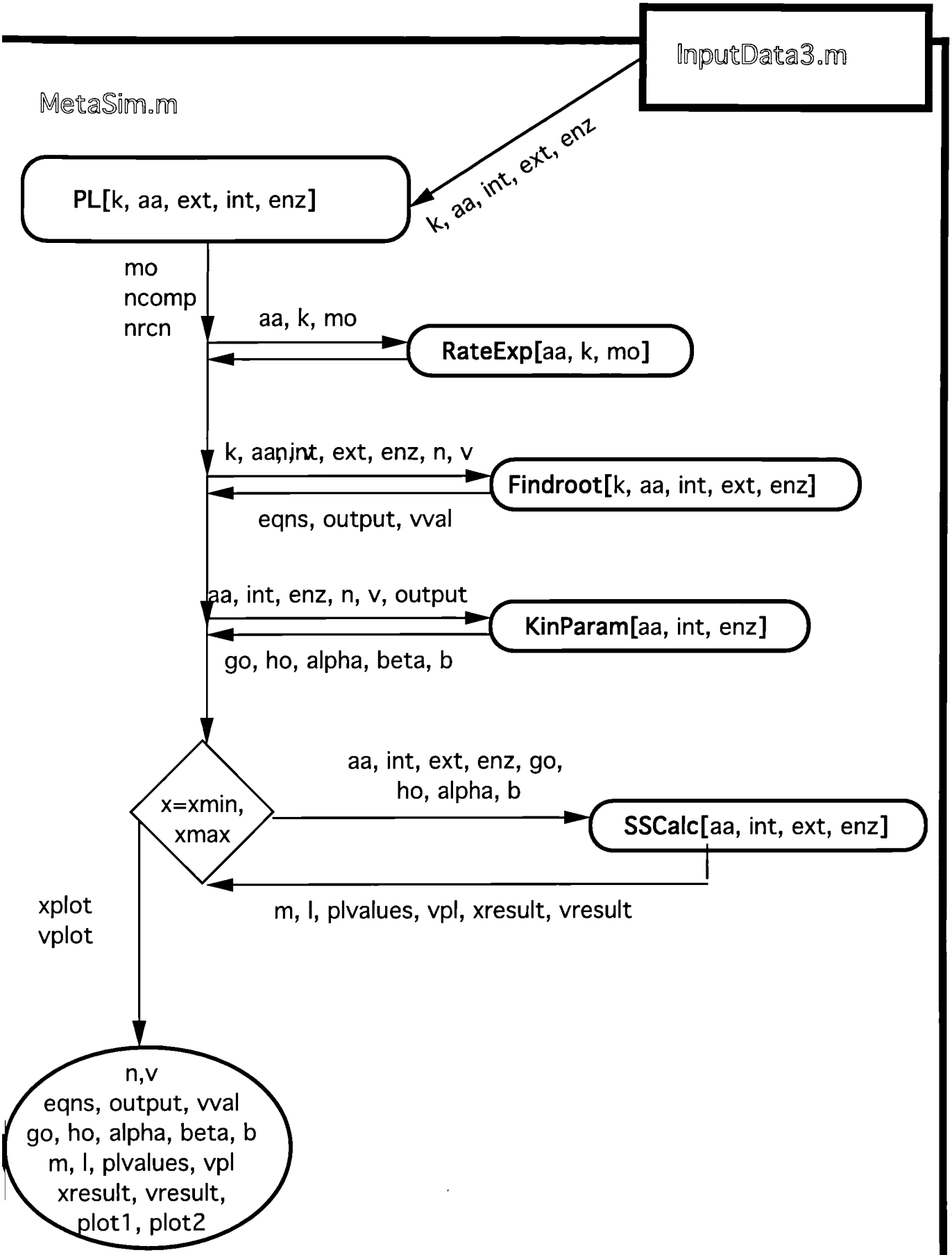
Noncompetitive Inhibition

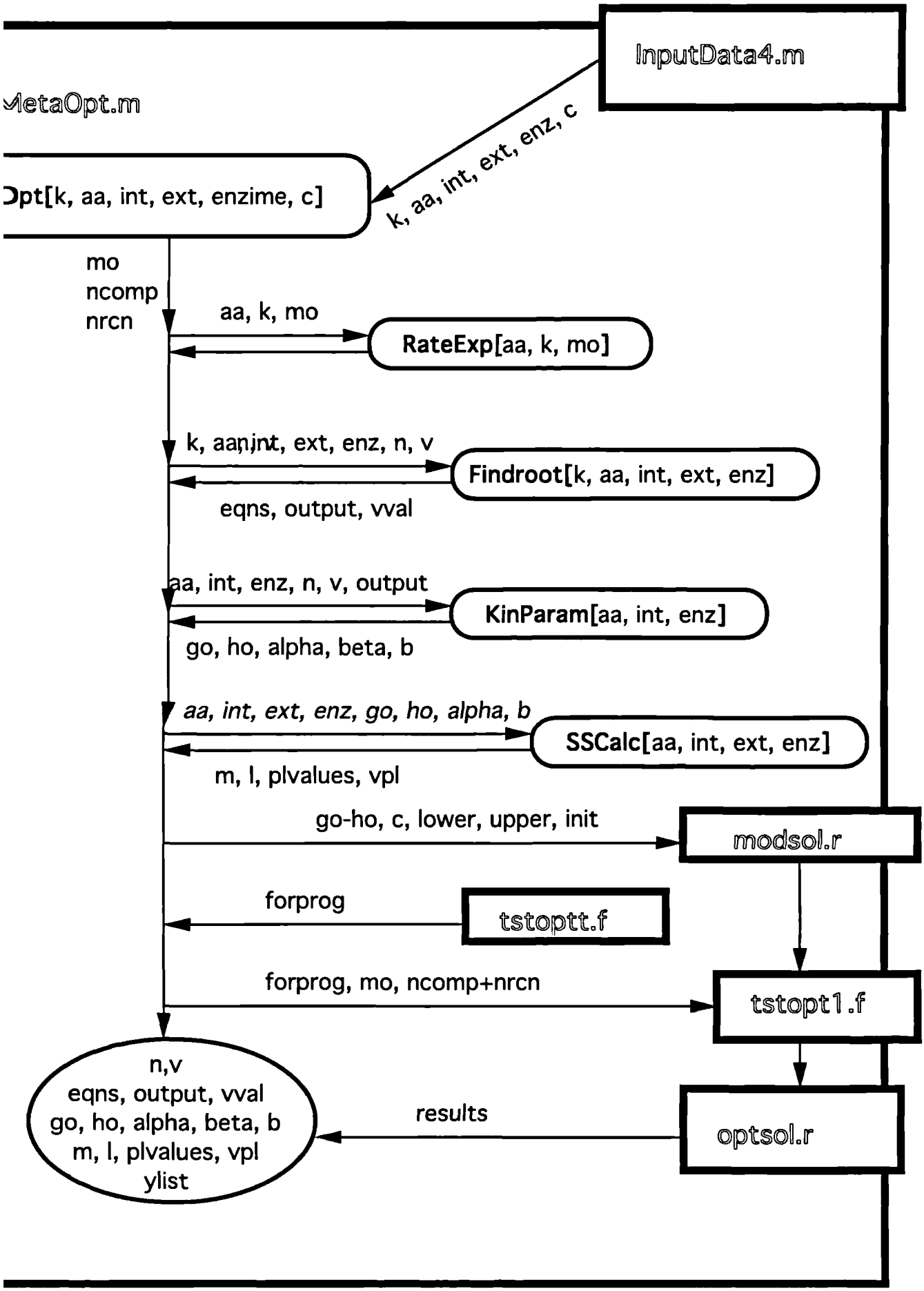
reversible reaction

$$C_{E_i}^{J} = \frac{1}{J}\left(\frac{\partial J}{\partial [I]}\right)_{I=0}\cdot -K_i\cdot \frac{\left(1 + \frac{[S]}{K_s} + \frac{[P]}{K_p}\right)}{\left(1 + \frac{[S]}{K_s}\right)}$$

irreversible reaction

$$C_{E_i}^{J} = \frac{1}{J}\left(\frac{\partial J}{\partial [I]}\right)_{I=0}\cdot -K_i$$


InputData3.m
MetaSim.m
k, aa, int, ext, enz
PL[k, aa, ext, int, enz]
mo
ncomp
nrcn
aa, k, mo
RateExp[aa, k, mo]
k, aa, int, ext, enz, n, v
Findroot[k, aa, int, ext, enz]
eqns, output, vval
aa, int, enz, n, v, output
KinParam[aa, int, enz]
go, ho, alpha, beta, b
x=xmin,
xmax
aa, int, ext, enz, go,
ho, alpha, b
SSCalc[aa, int, ext, enz]
m, l, plvalues, vpl, xresult, vresult
xplot
vplot
n,v
eqns, output, vval
go, ho, alpha, beta, b
m, l, plvalues, vpl
xresult, vresult,
plot1, plot2


InputData4.m
MetaOpt.m
Opt[k, aa, int, ext, enzime, c]
k, aa, int, ext, enz, c
mo
ncomp
nrcn
aa, k, mo
RateExp[aa, k, mo]
k, aa, int, ext, enz, n, v
Findroot[k, aa, int, ext, enz]
eqns, output, vval
aa, int, enz, n, v, output
KinParam[aa, int, enz]
go, ho, alpha, beta, b
aa, int, ext, enz, go, ho, alpha, b
SSCalc[aa, int, ext, enz]
m, l, plvalues, vpl
go-ho, c, lower, upper, init
modsol.r
forprog
tstoptt.f
forprog, mo, ncomp+nrcn
tstopt1.f
n,v
eqns, output, vval
go, ho, alpha, beta, b
m, l, plvalues, vpl
ylist
results
optsol.r

APPENDIX D

D.1 Instructions for using Meta on Sun Sparc:

1. Log in to user account: **machinename login:** "username"
password: "password"
2. Open windows environment: **1 machinename%** openwin
3. Open a cmdtool window by moving the cursor to an unused (grey) area of the screen, clicking on the right mouse button and choosing Command Tool from the Programs sub-menu.
4. In this window, we will log in to the machine suppporting Mathematica and set up the graphics connections as follows:

machinename % xhost bernadette
machinename % rlogin bernadette
1 bernadette % setenv DISPLAY machinename:0.0

5. To enter the Mathematica programme: 2 bernadette % math
6. There are demo programs which simulate the behaviour of the TOL meta-cleavage pathway which can be run.

To run a simulation of the TOL pathway on varying the input benzoate concentration:

In[1]:= <<Meta/Demo.m
In[2]:= <<Meta/SimDemo.m
In[3]:= PL[k, aa, ext, int, enzime]

To run an optimization of the TOL pathway:

In[1]:= <<Meta/Demo.m
In[2]:= <<Meta/OptDemo.m
In[3]:= Opt[k, aa, int, ext, enzime, c]

Descriptions of these subroutines, the input data required and the output parameters returned are given in the Figures D.1 and D.2 and in the following documentation.

Notation conventions: All references to functions or variables actually used in Mathematica are denoted by bold face.

D.2 Documentation of **MetaSim.m**:

Subroutine **PL**:

Purpose: This subroutine calculates the concentrations of the system parameters which result from varying one of the external metabolite concentrations as predicted by the power-law formalism. The calculations are made over a range, **xmin** to **xmax**, about an operating point **xmean**.

Parameters: **[k, aa, int, ext , enz]**

k: k is the matrix of kinetic parameters. Each row corresponds to the reaction of the same column number in the stoichiometric matrix. The kinetic parameters are normally those for one and two substrate Michaelis-Menten reactions, k_1 and K_m, and k_1, k_2, K_{12}, K_{21}, and K_m respectively. The choice of rate expression can however be changed by altering the expressions **f[x, e]** in **RateExp**.

aa: aa is the stoichiometric matrix, containing the relationships between the metabolites in the system. It has dimensions m x r+(m-m_o) where m is the total number of parameters, **r** is the number of reactions and **mo** is the number of dependent parameters. The coefficients of $[aa]_{ij}$ are as follows: $+\alpha$ if α moles of i are produced by reaction j, $-\alpha$ if α moles of i are consumed by reaction j, and zero if i does not take part in reaction j. The metabolites are numbered with the independent metabolites first, followed by the dependent metabolites.

int: int is the array of independent metabolite concentrations. It is entered in the form { {x[1], $conc_{X1}$},..., {x[i], $conc_{Ximean}$, $conc_{Ximin}$, $conc_{Ximax}$} ,..., {x[m-m_o], $conc_{Xm-mo}$} }.

ext: ext is the array of dependent metabolite concentrations. It is entered in the form { {x[m-m_o+1], $conc_{Xm-mo+1}$}, {x[m-m_o+2], $conc_{Xm-mo+2}$} ,..., {x[m], $conc_{Xm}$} }.

enz: **enz** is the array of enzyme concentrations. It is entered in the form { {e[1], $conc_{E1}$}, {e[2], $conc_{E2}$} ,..., {e[r], $conc_{Er}$} }.

Description:

1. The function **gi** assembles pairs of points for the graphical element **Line** from a list of data in the form X_i, Y_{1i}, Y_{2i},...,Y_{ni}.
2. The function **fi** attactes the label **x1** to the line of data points constructed from **list1**.
3. The values for the variable parameter **xi**, the operating point **xmean** and the range **xmin** to **xmax** are read from **ext**.
4. The number of dependent metabolites **mo** is calculated from the dimensions of **int**.
5. The number of components, **ncomp** and the number of reactions, **nrcn** are calculated from the stoichiometric matrix **aa**.
6. The arrays **xx** and **ee** which will contain metabolite and enzyme concentrations respectively are constructed.
7. **dep1**, which gives the value of the variable parameter **xi** at the operating point **xmean**, to be used in **KinParam** for calculating the power-law coefficents, is taken from **ext**.
5. These concentration arrays, **cint, cext** and **cenz** are constructed.
6. The range of x parameters which will be varied is calculated in **listxi**.
7. Headings for the concentration and flux data lists are constructed in **xresult** and **vresult** respectively.
8. The subroutine **Findroot** is called to calculate the steady state of the system using the kinetic parameters.
9. The subroutine **KinParam** is called to calculate the power-law parameters of the system.
10. The stream **str** is opened to the file **molsol.r**
11. The concentrations of the internal variables at the operating point **output**, the array of rates **v**, the matrix **go-ho** and **b** are written to the file.
12. The stream **str** is closed.
13. The iterative loop which runs through the values **xmin** to **xmax** is entered and **depi** is used instead of **ext** with the current value of **xi** inserted.
14. The subroutine **SSCalc** is called to calculate the power-law linear system.

15. The results arrays **vresult** and **xresult** are built up from the arrays **vpl** and **plvalues.**

16. These arrays are transformed into pairs of points **xplot** and **vplot** by the functions **gi** and **fi**, for plotting.

17. **plot1** and **plot2** are the graphical outputs of vplot and xplot respectively.

D.3 Documentation of **MetaOpt.m**:

Subroutine **Opt**

Purpose: This subroutine calculates the optimal set of system parameters for an objective function **c**. The system is approximated by a power-law model whose parameters are calculated from the kinetic parameters of the system. The system can then be expressed in linear terms so that linear optimization techniques can be applied. The linear optimization technique used is from a NAG subroutine **E04MBF**. This minimizes the objective function

$$f(x) = c^T x$$

subject to the equality and inequality constraints

$$\begin{aligned} A\,x &= b \\ x &< b_1 \\ x &> b_2 \end{aligned}$$

These constraints are collected together into an upper bound and a lower bound so that

$$b_{lower} < A^t x < b_{upper}$$

Parameters: **[k, aa, int, ext , enz, c]**

k: k is the matrix of kinetic parameters. Each row corresponds to the reaction of the same column number in the stoichiometric matrix. The kinetic parameters are normally those for one and two substrate Michaelis-Menten reactions, k_1 and K_m, and k_1, k_2, K_{12}, K_{21}, and K_m respectively. The choice of rate expression can however be changed by altering the expressions **f[x, e]** in **RateExp.**

aa: aa is the stoichiometric matrix, containing the relationships between the metabolites in the system. It has dimensions m x r+(m-m_o) where m is the total number of parameters, **r** is the number of reactions and **mo** is the number of dependent parameters. The coefficients of $[aa]_{ij}$ are as follows: $+\alpha$ if α moles of i are produced by reaction j, $-\alpha$ if α moles of i are consumed by reaction j, and zero if i does not take part in reaction j. The metabolites are numbered with the independent metabolites first, followed by the dependent metabolites.

int: int is the array of independent metabolite concentrations. It is entered in the form { {x[1], $conc_{X1}$}, {x[2], $conc_{X2}$} ,..., {x[m-m_o], $conc_{Xm-mo}$} }.

ext: ext is the array of dependent metabolite concentrations. It is entered in the form { {x[m-m_o+1], $conc_{Xm-mo+1}$}, {x[m-m_o+2], $conc_{Xm-mo+2}$} ,..., {x[m], $conc_{Xm}$} }.

enz: enz is the array of enzyme concentrations. It is entered in the form { {e[1], $conc_{E1}$}, {e[2], $conc_{E2}$} ,..., {e[r], $conc_{Er}$} }.

c: c is the objective function of the optimization problem as defined in Eqn.D1 above. For example, if it is required to minimize X_p then c_i=0, i n.e. p, c_i=1, i=p.

Description:

1. The number of dependent metabolites **mo** is calculated from the dimensions of **int.**
2. The number of components, **ncomp** and the number of reactions, **nrcn** are calculated

from the stoichiometric matrix **aa**.

3. The arrays **xx** and **ee** which will contain metabolite and enzyme concentrations respectively are constructed.

4. **dep1**, which is used in **KinParam** is taken from **ext**.

5. These concentration arrays, **cint**, **cext** and **cenz** are constructed.

6. The subroutine **Findroot** is called to calculate the steady state of the system using the kinetic parameters.

7. The subroutine **KinParam** is called to calculate the power-law parameters of the system.

8. The subroutine **SSCalc** is called to calculate the power-law linear system.

9. The initial guess for the numerical algorithm **init** is the (feasible) operating point.

10. The matrix of the equality contraints, **aout**, is calculated from the power-law parameters **go** and **ho**.

11. **lower** is the array of the lower bounds.

12. **upper** is the array of the upper bounds.

The next six lines of code open a data file called **modelsol.r** and write into it the data required for the optimization algorithm

13. The stream **str** is opened to the file **molsol.r**

14. The objective function **c**, the matrix **aout**, the upper and lower bounds **bupper** and **blower** and the initial guess **init** are written to the file.

15. The stream **str** is closed.

The next five lines of code read in the Fortran program which calls the numerical algorithm **E04MBF** and enter the dimensions of the problem.

16. The Fortran program **tstopt.f** is read in as a list of characters, **forprog**.

17. The dimensions of the system **mo** and **ncomp+nrcn** are entered in the appropriate place in the list at positions 200 and 216.

18. A stream to the file **tstopt1.f**, **str1** is opened.

19. All of the characters in the amended list **forprog** are joined together and written to the file **tstopt1.f** via the stream **str1**, which is then closed.

20. This file is then compiled on the f77 compiler

21. The executable file **a.out** is run.

The next four lines read in the optimization results from the file **optsol.r**

22. The results file **optsol.r** is read as a list of word-delineated strings into the array results.
23.The enzyme and intermediate concentration data required is in the fourth column in positions 6 to 5+nrcn+ncomp and this is pulled out into **xstring.**
24. This string is then converted into a list of numbers, **ylist,** and plotted as a barchart.

Subroutine **Findroot**

Purpose: The subroutine **Findroot** calculates the steady state of a system of metabolic equations. The values of the dependent variables are calculated from the rate expressions and the values of the independent variables. The set of reactions must be linearly independent and the number of independent reactions must equal the number of dependent variables so that the system is not under- or over-determined.

Parameters: **[k, aa, int, ext , enz]**

k: k is the matrix of kinetic parameters. Each row corresponds to the reaction of the same column number in the stoichiometric matrix. The kinetic parameters are normally those for one and two substrate Michaelis-Menten reactions, k_1 and K_m, and k_1, k_2, K_{12}, K_{21}, and K_m respectively. The choice of rate expression can however be changed by altering the expressions **f[x, e]** in **RateExp.**

aa: aa is the stoichiometric matrix, containing the relationships between the metabolites in the system. It has dimensions m x r+(m-m_o) where m is the total number of parameters, **r** is the number of reactions and **mo** is the number of dependent parameters. The coefficients of $[aa]_{ij}$ are as follows: $+\alpha$ if α moles of i are produced

by reaction j, $-\alpha$ if α moles of i are consumed by reaction j, and zero if i does not take part in reaction j. The metabolites are numbered with the independent metabolites first, followed by the dependent metabolites.

int: int is the array of independent metabolite concentrations. It is entered in the form { {x[1], $conc_{X1}$}, {x[2], $conc_{X2}$} ,..., {x[m-m_o], $conc_{Xm-mo}$} }.

ext: ext is the array of dependent metabolite concentrations. It is entered in the form { {x[m-m_o+1], $conc_{Xm-mo+1}$}, {x[m-m_o+2], $conc_{Xm-mo+2}$} ,..., {x[m], $conc_{Xm}$} }.

enz: enz is the array of enzyme concentrations. It is entered in the form { {e[1], $conc_{E1}$}, {e[2], $conc_{E2}$} ,..., {e[r], $conc_{Er}$} }.

Description:

1. The number of dependent metabolites **mo** is calculated from the dimensions of $_{int.}$
2. The arrays **xx** and **ee** which will contain metabolite and enzyme concentrations respectively are constructed.
3. These concentration arrays, **cint, cext** and **cenz** are constructed.
4. The subroutine **RateExp** is called to calculate the reduced stoichiometric matrix **n** and the array of reaction rate expressions **v**.
5. The system steady state balance equations **eqns** are calculated from **n.v**.
6. The rule **eval** is defined. This enables x[i] to be evaluated as $conc_{Xi}$.
7. Applying this rule to **cext** and **cenz** evaluates x[i] and e[i] numerically as $conc_{Xi}$ and $conc_{Ei}$ in all following expressions.
8. The rule **hi** enables the Mathematica subroutine **FindRoot** to be applied to a list of initial guesses.
9. Applying this rule and the subroutine **FindRoot** to **eqns** and the initial guesses **cint** yields the steady state solution in the form { x[m-mo+1] -> $conc_{Xm-mo+1}$,..., x[m] -> $conc_{Xm}$ }
10. The values of the reaction rates are obtained by using these values in **v** to obtain

vval.

11. Clearing **e** and **x** removes the numerical value and allows them to be evaluated symbolically.

Subroutine **KinParam**

Purpose: The purpose of this subroutine is to calculate the power-law parameters of a system from a steady state operating point and the system rate expressions. The power-law coefficients describe the behaviour of the system about the operating point according to the equation

$$dX_i/dt = V_i - V_{-i}$$
$$= \alpha_i \prod_{j=1}^{n} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n} X_j^{h_{ij}} \qquad i = 1,\ldots,n.$$

g_{ij} and h_{ij} are evaluated from the following equations:

$$g_{ij} = \partial(\ln v_{iO})/\partial(\ln X_j) = (\partial v_{iO}/\partial X_j)(X_{jO}/v_{iO})$$

$$h_{ij} = \partial(\ln v_{-iO})/\partial(\ln X_j) = (\partial v_{-iO}/\partial X_j)(X_{jO}/v_{-iO})$$

and the parameters α_i and β_i from

$$\alpha_i = v_{iO} \prod_{j=1}^{n} X_{jO}^{-g_{ij}}$$

$$\beta_i = v_{-iO} \prod_{j=1}^{n} X_{jO}^{-h_{ij}}$$

Parameters:

aa: aa is the stoichiometric matrix, containing the relationships between the metabolites in the system. It has dimensions m x r+(m-m_o) where m is the total number of parameters, **r** is the number of reactions and **mo** is the number of dependent parameters. The coefficients of $[aa]_{ij}$ are as follows: $+\alpha$ if α moles of i are produced by reaction j, $-\alpha$ if α moles of i are consumed by reaction j, and zero if i does not take part in reaction j. The metabolites are numbered with the independent metabolites first, followed by the dependent metabolites.

int: int is the array of independent metabolite concentrations. It is entered in the form { {x[1], $conc_{X1}$}, {x[2], $conc_{X2}$} ,..., {x[m-m_o], $conc_{Xm-mo}$} }.

ext: ext is the array of dependent metabolite concentrations. It is entered in the form { {x[m-m_o+1], $conc_{Xm-mo+1}$}, {x[m-m_o+2], $conc_{Xm-mo+2}$} ,..., {x[m], $conc_{Xm}$} }.

enz: enz is the array of enzyme concentrations. It is entered in the form { {e[1], $conc_{E1}$}, {e[2], $conc_{E2}$} ,..., {e[r], $conc_{Er}$} }.

The following two input parameters are calculated by RateExp

n: n is the reduced stoichiometric matrix.

v: v is the array of rate expressions corresponding to the **r** system reactions.

output: output is the set of steady state values of the dependent variables. **output** is calculated by the subroutine **Findroot**.

Description:

1. The rule **vform** uses the reduced stoichiometric matrix **n** with the array of reaction rate expressions to calculate the arrays of aggregated fluxes V_i and V_{-i}. V_i is the sum of the fluxes which go to increase the pool of X_i and V_{-i} is the sum of the fluxes which go to decrease the pool of X_i.

2.. The number of dependent metabolites **mo** is calculated from the dimensions of **int**.

3. The arrays **xx** and **eee** which will contain metabolite and enzyme concentrations respectively are constructed.

4. These concentration arrays, **cint, cext** and **cenz** are constructed.

5. **xall** constructs a full list of the system parameters, which includes independent metabolites, dependent metabolites, and enzymes. The enzymes are renamed as x[m+1] ,.., x[m+r] for ease in numerical computation.

6. **vminus** is V_{-i}, i=1 ,..., m_o, the array of aggregated fluxes which go to decrease the pool of X_i. .

7. vplus is V_i, i=1 ,..., m_o the array of aggregated fluxes which go to increase the pool of X_i.

8. **g** is the matrix of power-law coefficients calculated from Eqn.D5 above.

9. **h** is the matrix of power-law coefficients calculated from Eqn.D6 above.

The next 6 lines insert numerical values for those of **x[i]** and **e[i]** in the expressions for **g, h, vplus, vminus** and **xall.**

10. Numerical values for the independent metabolite and enzymes concentrations are entered in all expressions.

11. Numerical values for dependent metabolite concentrations are entered in **g.**

12. Numerical values for dependent metabolite concentrations are entered in **h.**

13. Numerical values for dependent metabolite concentrations are entered in **vplus.**

14. Numerical values for dependent metabolite concentrations are entered in **vminus.**

15. The power-law parameters α_i (**alpha**) are calculated from Eqn.D7 above.

16. The power-law parameters β_i (**beta**) are calculated from Eqn.D8 above.

17. The output array **b** = α/β is calculated.

18. Clearing all values of **x** removes the numerical values and allows them to be evaluated symbolically.

Subroutine SSCalc

Purpose: The subroutine **SSCalc** calculates values for the steady state that is reached when the system is perturbed from its original operating point, using a power-law model and coefficients as its model. The system is linear in logarithmic coordinates and can be described by the equation

$$Ay = b$$

where $A = g_{ij} - h_{ij}$, $b = \alpha_i/\beta_i$ and $y = \text{Ln } x_i$. The solution is obtained by partitioning the matrix A into an m_o x m_o matrix **a1** and an m_o x $m\text{-}m_o+r$ matrix **a2** and the vector of system parameters y into [y':y'']. The solution is of the form

$$y' = mb + ly''$$

where $m = [\mathbf{a1}]^{-1}$ and $l = -[\mathbf{a1}]^{-1}\mathbf{a2}$. Both metabolite concentrations in logarithmic form (**ypl**) and rates (**vpl**) are calculated.

Parameters:

aa: aa is the stoichiometric matrix, containing the relationships between the metabolites in the system. It has dimensions m x $r+(m\text{-}m_o)$ where m is the total number of parameters, **r** is the number of reactions and **mo** is the number of dependent parameters. The coefficients of $[aa]_{ij}$ are as follows: $+\alpha$ if α moles of i are produced by reaction j, $-\alpha$ if α moles of i are consumed by reaction j, and zero if i does not take part in reaction j. The metabolites are numbered with the independent metabolites first, followed by the dependent metabolites.

int: int is the array of independent metabolite concentrations. It is entered in the form $\{ \{x[1], conc_{X1}\}, \{x[2], conc_{X2}\}, \ldots, \{x[m\text{-}m_o], conc_{Xm\text{-}mo}\} \}$.

ext: ext is the array of dependent metabolite concentrations. It is entered in the form { {x[m-m_o+1], $conc_{Xm-mo+1}$}, {x[m-m_o+2], $conc_{Xm-mo+2}$} ,..., {x[m], $conc_{Xm}$} }.

enz: enz is the array of enzyme concentrations. It is entered in the form { {e[1], $conc_{E1}$}, {e[2], $conc_{E2}$} ,..., {e[r], $conc_{Er}$} }.

The following two input parameters are calculated by **RateExp**

n: n is the reduced stoichiometric matrix.
v: v is the array of rate expressions corresponding to the **r** system reactions.

output: output is the set of steady state values of the dependent variables. **output** is calculated by the subroutine **Findroot**.

go: go is the m_o x m+r matrix of power-law coefficients defined by Eqn.D5 and used in the linear expression above to model the system.
ho: ho is the m_o x m+r matrix of power-law coefficients defined by Eqn.D6 and used in the linear expression above to model the system.
b: b=alpha/beta where **alpha** and **beta** are the power-law coefficients defined by Eqn.s D7 and D8 and used in the linear expression above to model the system.
go, ho and **b** are calculated **by KinParam.**

Description:

1. The number of dependent metabolites **mo** is calculated from the dimensions of **int.**
2. The arrays **xx** and **eee** which will contain metabolite and enzyme concentrations respectively are constructed.
3. These concentration arrays, **cint, cext** and **cenz** are constructed.
4. The matrix **a = go - ho** to be used in the linear description of the system Ay = b is calculated.
5 and 6. The m_o x r matrix **a** is partitioned into an m_o x m_o matrix **a1** and an m_o x m-m_o+r matrix **a2.**

7. **x2** is the list of independent parameters, including the enzyme concentrations.

8. **y2** is the natural log of **x2**.

9. Applying this rule to **cext** and **cenz** evaluates x[i] and e[i] numerically as $conc_{Xi}$ and $conc_{Ei}$ in all following expressions.

10. The matrices **m** and **l** defined in Eqn.D10 above are calculated.

11. **plvalues**, the array of logarithmic values of the internal system parameters as predicted by a power-law model, is calculated.

12. **ypl**, the array of logarithmic values of all of the system parameters as predicted by a power-law model, is calculated.

13. **vpl**, the array of logarithmic values of the system fluxes as predicted by a power-law model, is calculated.

14. Clearing all values of **x** removes the numerical values and allows them to be evaluated symbolically.

D.4 Documentation of **InVivo.m**

Purpose: This subroutine calculates the flux control coefficients for a metabolic pathway based upon transient metabolite concentrations. This is done by calculating correlation coefficients α in the following expression:

$$\sum_{i=1}^{n} \alpha_i \left(x_i[t=t_j] - x_i[t=t_0] \right) = t_j \qquad j = 1,\ldots,p$$

The flux control coefficients can then be calculated from the relationship:

$$C = \alpha A J$$

This program uses a NAG subroutine C04NCF for the regression analysis.

Parameters:

aa: aa is the stoichiometric matrix, containing the relationships between the metabolites in the system. It has dimensions m x r+(m-m_o) where m is the total number of parameters, **r** is the number of reactions and **mo** is the number of dependent parameters. The coefficients of $[aa]_{ij}$ are as follows: $+\alpha$ if α moles of i are produced by reaction j, $-\alpha$ if α moles of i are consumed by reaction j, and zero if i does not take part in reaction j. The metabolites are numbered with the independent metabolites first, followed by the dependent metabolites.

mo: mo is the number of dependent metabolites.

Description:

1. The number of components **ncomp** is calculated from the dimensions of the stoichiometric matrix **aa**.
2. The reduced stoichiometric matrix **a (ncomp x nrcn)** is extracted from **aa.**
3. The transient concentration data is read in as a series of columns of time and concentration data.
4. The columns for time and benzoic acid concentration are extracted as **xstring1** and **xstring2** and are converted into the expressions **time** and **benz** respectively.

The next five lines of code read in the Fortran program which calls the numerical algorithm **C04NCF** and enter the dimensions of the problem.

5. The Fortran program **invivot.f** is read in as a list of characters, **forprog.**
6. The dimensions of the system **ncomp** and the number of time parameters are entered in the appropriate place in the list at positions 164 and 171.
7. A stream to the file **invivo1.f, str1** is opened.
8. All of the characters in the amended list **forprog** are joined together and written to the file **invivo1.f** via the stream **str1**, which is then closed.
9. This file is then compiled on the f77 compiler
10. The executable file **a.out** is run.

The next four lines read in the optimization results from the file **invivo.r**

11. The results file **invivo.r** is read as a list of word-delineated strings into the array results.

12. The corrlation coefficients required are in the second column in positions 7 to 6+ncomp and this is pulled out into **xstring**.

13. This string is then converted into a list of numbers, **alpha**.

14. The system flux **ji** is calculated from the changing benzoate concentration and an **nrcn x nrcn** diagonal matrix **j** = [ji] is constructed.

15. The set of correlation coefficients **coeffs** is calculated from the expression D12 above and plotted as a barchart.

D.5 Program Listings

MetaSim.m

```
BeginPackage["ModelSolve3`"]


RateExp::usage = "does rate expressions"

Findroot::usage = "finds roots"

KinParam::usage = "calculates BST kinetic parameters"

SSCalc::usage = "calculate steady state parameters"

PL::usage = "plots changing s.s. with changing input parameters"

Begin["`Private`"]

RateExp[aa_, k_, mo_] :=
        Block[{i, ncomp, nrcn, f, enzyme, nl1},
                ncomp = Length[aa];
                nrcn = Length[Transpose[aa]] - (ncomp - mo);
                xx = Table[x[i], {i, Length[aa]}];
                ee = Table[e[i], {i, nrcn}];
                pos[x_] := 0 /; x > 0;
                pos[x_] := -x /; x <= 0;
                r = Map[ Drop[# , -(ncomp-mo)]&, aa];
                n = Drop[r, (ncomp-mo)];
                f[{k1_, k2_}, {x_}, {e_}] := k1 e x / ( k2 + x );
                f[{k1_, k2_, k3_, k4_, k5_}, {x1_, x2_}, {e_}] := k1 e /
                        (1 + k5/x1 + k4/x2 + (k5 k3 + k2 k4)/ (2 x1 x2) );
```

```
            nll = Map[List, Map[pos, Transpose[r], {2}].xx] /. List[Plus[x_, y_]]
-> List[x, y];(*note: try List[Plus[z_]] -> List[z] *)
            enzyme = Map[List, e];
            v = Thread[ f[k, nll, enzyme] ];
      ]
Findroot[k_, aa_, int_, ext_, enz_] :=
      Block[{hi, z },
            mo = Length[int];
            xx = Table[x[i], {i, Length[aa]}];
            ee = Table[e[i], {i, Length[Transpose[aa]]-(Length[aa]-mo)}];
            cenz = Transpose[{ee, Transpose[enz][[2]]}];
            cext = Transpose[{Take[xx, Length[ext]], Transpose[ext][[2]]}];
            cint = Transpose[{Take[xx, -Length[int]], Transpose[int][[2]]}];
            xext = Transpose[cext][[1]];
            RateExp[aa, k, mo];
            eqns = n.v;
            eval[z_] := Map[Apply[Set, #]&, z];
            hi[eqn__, {lin__}] := FindRoot[eqn, lin];
            eval[cext];eval[cenz];
            output = hi[eqns, cint];
            vval = v/.output;
            Clear[x];Clear[e];
      ]

KinParam[aa_, int_, ext_, enz_] :=
      Block[{z },
            vform[r_, vv_, e_] := Transpose[Map[pos, Transpose[r], {-1}]].vv/.
e[z_] -> x[z+Length[aa]];
            mo = Length[int];
            xx = Table[x[i], {i, Length[aa]}];
            eee = Table[x[i], {i, Length[aa]+1, Length[Transpose[aa]] + mo}];
            cenz = Transpose[{eee, Transpose[enz][[2]]}];
```

```
            cext = Transpose[{Take[xx, Length[ext]], Transpose[ext][[2]]}];
            cint = Transpose[{Take[xx, -Length[int]], Transpose[int][[2]]}];
            xall = Union[Transpose[cext][[1]], Transpose[cint][[1]],
Transpose[cenz][[1]]];
            vminus = vform[n, v, e];
            vplus = vform[-n, v, e];
            g = Transpose[ Table[ (x[i]/vplus[[j]]) D[ vplus[[j]], x[i] ], {i,
Length[aa]+Length[v]}, {j, mo} ] ];
            h = Transpose[ Table[ (x[i]/vminus[[j]]) D[ vminus[[j]], x[i] ], {i,
Length[aa]+Length[v]}, {j, mo} ] ];
            eval[cext]; eval[cenz];
            go = g /. output;
            ho = h /. output;
            xo = xall /. output;
            vpluso = vplus /. output;
            vminuso = vminus /. output;
            alpha = Table [ vpluso[[j]] Product[ x[i]^-go[[j,i]], {i, Length[xall]} ],
{j, mo}]/.output ;
            beta = Table [ vminuso[[j]] Product[ x[i]^-ho[[j,i]], {i, Length[xall]} ],
{j, mo}]/.output;
            b = Log[ beta/alpha ];
            Clear[x]
      ]

SSCalc[aa_, int_, ext_, enz_] :=
      Block[{a2, i },
            mo = Length[int];
            xx = Table[x[i], {i, Length[aa]}];
            eee = Table[x[i], {i, Length[aa]+1, Length[Transpose[aa]] + mo}];
            cenz = Transpose[{eee, Transpose[enz][[2]]}];
            cext = Transpose[{Take[xx, Length[ext]], Transpose[ext][[2]]}];
            cint = Transpose[{Take[xx, -Length[int]], Transpose[int][[2]]}];
```

```
            a = go - ho;
            a1 = Transpose[Table[Transpose[a][[i]], {i, Length[aa]-mo+1,
Length[aa]}]];
            a2 = Transpose[Join[Table[Transpose[a][[i]], {i, Length[aa]-mo}],
Table[Transpose[a][[i]], {i, Length[aa]+1, Length[xall]}]]];
            x2 = Join[ Table[ xall[[i]], {i, Length[aa]-mo} ], Table[ xall[[i]], {i,
Length[aa]+1, Length[xall]} ] ];
            y2 = Log[x2];
            eval[cext];eval[cenz];
            m = Inverse[a1]; l = -Inverse[a1].a2;
            plvalues = m.b + l.y2;
            temp = Transpose[cint];
            temp[[2]] = Exp[plvalues];
            eval[Transpose[temp]];
            ypl = Log[xall];
            vpl = Log[alpha] + go.ypl;
            Clear[x]
      ]

PL[k_, aa_, ext_?(MemberQ[#, {xi_, xmean_, xmin_, xmax_}]&), int_, enz_] :=
      Block[{i },
            gi[xindepi_, listxi_] := Line[Thread[List[listxi, Drop[xindepi, 1]]]];
            fi[Line[list1_], x1_] :={Line[list1](*, Text[x1, Last[list1], {0, -1}]*)};
            {xi, xmean, xmin, xmax} = Flatten[Cases[ext, {x1_, x2_, x3_, x4_}],
1];
            mo = Length[int];
            ncomp = Length[aa];
            nrcn = Length[Transpose[aa]] - (ncomp-mo);
            xx = Table[x[i], {i, Length[aa]}];
            ee = Table[e[i], {i, Length[Transpose[aa]]-(Length[aa]-mo)}];
            dep1 = ext/.{xi, xmean, xmin, xmax} -> {xi, xmean};
            cenz = Transpose[{ee, Transpose[enz][[2]]}];
```

```
        cext = Transpose[{Take[xx, Length[dep1]], Transpose[dep1][[2]]}];
        cint = Transpose[{Take[xx, -Length[int]], Transpose[int][[2]]}];
        listxi = Table[i, {i, xmin, xmax}];
        xresult = {Transpose[int][[1]]};
        vlist = Table[rateV[i], {i, mo}];
        vresult = {vlist};
        Findroot[k, aa, int, dep1, enz];
        KinParam[aa, int, dep1, enz];
        str = OpenWrite["modsol.r"];
        Write[str, output, v, go-ho, b]; Close[str];
        Do[   depi = ext/.{xi, xmean, xmin, xmax} -> {xi, listxi[[j]]};
              SSCalc[aa, int, depi, enz];
              vresult = Append[vresult, vpl];
              xresult = Append[xresult, plvalues],
          {j, Length[listxi]} ];
        xplot = Thread[fi[Map[gi[#, listxi]&, N[Transpose[xresult]]],
Log[Transpose[int][[1]]]]];
        vplot = Thread[fi[Map[gi[#, listxi]&, N[Transpose[vresult]]], vlist]];
 plot2 = Show[Graphics[xplot], Axes -> Automatic, AxesLabel -> {FontForm[
"Conc.X[1] (mM)", {"Times", 10}], FontForm["Log X[i] (mM)", {"Times", 10}] },
AxesOrigin -> {20.0, 0.0}, DefaultFont -> {"Times-Italic", 8}];
plot1 = Show[Graphics[vplot], Axes -> Automatic, AxesLabel -> {FontForm[
"Conc.X[1] (mM)",  {"Times", 10}], FontForm["Flux V[i]", {"Times", 10}] },
AxesOrigin -> {20.0, 5.85}, DefaultFont -> {"Times-Italic", 8}];
Show[GraphicsArray[{{plot2}, {plot1}}]]
    ]

(*Opt[k_, aa_, ext_, int_, enz_, copt_] :=
    Findroot[k, aa, int, ext, enz];
    KinParam[aa, int, ext, enz];
     SSCalc[aa, int, ext, enz];
    str = OpenWrite["opt.d"];
```

```
    Write[str, copt, Prepend[Flatten[go-ho], bit], Append[b, bit]];
    Close[str];*)


End[ ]
EndPackage[]
```

MetaOpt.m

```
BeginPackage["Opt`", "Graphics`Graphics`"]

Opt::usage = "finds optimal set of system parameters for a given objective function."

RateExp::usage = "does rate expressions"

Findroot::usage = "finds roots"

KinParam::usage = "calculates BST kinetic parameters"

SSCalc::usage = "calculate steady state parameters"
Begin["`Private`"]

RateExp[aa_, k_, mo_] :=
        Block[{i, ncomp, nrcn, f, enzyme, nl1},
                ncomp = Length[aa];
                nrcn = Length[Transpose[aa]] - (ncomp - mo);
                xx = Table[x[i], {i, Length[aa]}];
                ee = Table[e[i], {i, nrcn}];
                pos[x_] := 0 /; x > 0;
                pos[x_] := -x /; x <= 0;
                r = Map[ Drop[# , -(ncomp-mo)]&, aa];
                n = Drop[r, (ncomp-mo)];
                f[{k1_, k2_}, {x_}, {e_}] := k1 e x / ( k2 + x );
                f[{k1_, k2_, k3_, k4_, k5_}, {x1_, x2_}, {e_}] := k1 e /
                        (1 + k5/x1 + k4/x2 + (k5 k3 + k2 k4)/ (2 x1 x2) );
                nl1 = Map[List, Map[pos, Transpose[r], {2}].xx] /. List[Plus[x_, y_]]
-> List[x, y];(*note: try List[Plus[z__]] -> List[z] *)
                enzyme = Map[List, ee];
```

```
            v = Thread[ f[k, nl1, enzyme] ];
        ]

Findroot[k_, aa_, int_, ext_, enz_] :=
        Block[{hi, z },
            mo = Length[int];
            xx = Table[x[i], {i, Length[aa]}];
            ee = Table[e[i], {i, Length[Transpose[aa]]-(Length[aa]-mo)}];
            cenz = Transpose[{ee, Transpose[enz][[2]]}];
            cext = Transpose[{Take[xx, Length[ext]], Transpose[ext][[2]]}];
            cint = Transpose[{Take[xx, -Length[int]], Transpose[int][[2]]}];
            xext = Transpose[cext][[1]];
            RateExp[aa, k, mo];
            eqns = n.v;
            eval[z_] := Map[Apply[Set, #]&, z];
            hi[eqn__, {lin__}] := FindRoot[eqn, lin];
            eval[cext];eval[cenz];
            output = hi[eqns, cint];
            vval = v/.output;
            Clear[x];Clear[e];
        ]

KinParam[aa_, int_, ext_, enz_] :=
        Block[{z },
            vform[r_, vv_, e_] := Transpose[Map[pos, Transpose[r], {-1}]].vv/.
e[z_] -> x[z+Length[aa]];
            mo = Length[int];
            xx = Table[x[i], {i, Length[aa]}];
            eee = Table[x[i], {i, Length[aa]+1, Length[Transpose[aa]] + mo}];
            cenz = Transpose[{eee, Transpose[enz][[2]]}];
            cext = Transpose[{Take[xx, Length[ext]], Transpose[ext][[2]]}];
            cint = Transpose[{Take[xx, -Length[int]], Transpose[int][[2]]}];
```

```
            xall = Union[Transpose[cext][[1]], Transpose[cint][[1]],
Transpose[cenz][[1]]];
            vminus = vform[n, v, e];
            vplus = vform[-n, v, e];
            g = Transpose[ Table[ (x[i]/vplus[[j]]) D[ vplus[[j]], x[i] ], {i,
Length[aa]+Length[v]}, {j, mo} ] ];
            h = Transpose[ Table[ (x[i]/vminus[[j]]) D[ vminus[[j]], x[i] ], {i,
Length[aa]+Length[v]}, {j, mo} ] ];
            eval[cext]; eval[cenz];
            go = g /. output;
            ho = h /. output;
            xo = xall /. output;
            vpluso = vplus /. output;
            vminuso = vminus /. output;
            alpha = Table [ vpluso[[j]] Product[ x[i]^-go[[j,i]], {i, Length[xall]} ],
{j, mo}]/.output ;
            beta = Table [ vminuso[[j]] Product[ x[i]^-ho[[j,i]], {i, Length[xall]} ],
{j, mo}]/.output;
            b = Log[ beta/alpha ];
            Clear[x]
        ]

SSCalc[aa_, int_, ext_, enz_] :=
        Block[{a2, i },
            mo = Length[int];
            xx = Table[x[i], {i, Length[aa]}];
            eee = Table[x[i], {i, Length[aa]+1, Length[Transpose[aa]] + mo}];
            cenz = Transpose[{eee, Transpose[enz][[2]]}];
            cext = Transpose[{Take[xx, Length[ext]], Transpose[ext][[2]]}];
            cint = Transpose[{Take[xx, -Length[int]], Transpose[int][[2]]}];
            a = go - ho;
            a1 = Transpose[Table[Transpose[a][[i]], {i, Length[aa]-mo+1,
```

```
Length[aa]}]];
            a2 = Transpose[Join[Table[Transpose[a][[i]], {i, Length[aa]-mo}],
Table[Transpose[a][[i]], {i, Length[aa]+1, Length[xall]}]]];
            x2 = Join[ Table[ xall[[i]], {i, Length[aa]-mo} ], Table[ xall[[i]], {i,
Length[aa]+1, Length[xall]} ] ];
            y2 = Log[x2];
            eval[cext];eval[cenz];
            m = Inverse[a1]; l = -Inverse[a1].a2;
            plvalues = m.b + l.y2;
            temp = Transpose[cint];
            temp[[2]] = Exp[plvalues];
            eval[Transpose[temp]];
            ypl = Log[xall];
            vpl = Log[alpha] + go.ypl;
            Clear[x]
      ]

Opt[k_, aa_, int_, ext_, enz_, c_] :=
      Block[{i },
            mo = Length[int];
            ncomp = Length[aa];
            nrcn = Length[Transpose[aa]] - (ncomp-mo);
            xx = Table[x[i], {i, Length[aa]}];
            ee = Table[e[i], {i, Length[Transpose[aa]]-(Length[aa]-mo)}];
            dep1 = ext;
            cenz = Transpose[{ee, Transpose[enz][[2]]}];
            cext = Transpose[{Take[xx, Length[dep1]], Transpose[dep1][[2]]}];
            cint = Transpose[{Take[xx, -Length[int]], Transpose[int][[2]]}];
            Findroot[k, aa, int, dep1, enz];
            KinParam[aa, int, dep1, enz];
            SSCalc[aa, int, dep1, enz];
            init = Join[ Log[Transpose[cext][[2]]], plvalues,
```

```
Log[Transpose[cenz][[2]]] ];
                aout = Flatten[go - ho];
                lower = Table[-100., {i, ncomp+nrcn}];
                upper = Table[10., {i, ncomp+nrcn}];
                str = OpenWrite["modsol.r"];
                Write[str, Prepend[Append[c, s], s]];
                Write[str, Prepend[Append[aout, s], s]];
                Write[str, Prepend[Append[Join[lower, b], s], s]];
                Write[str, Prepend[Append[Join[upper, b], s], s]];
                Write[str, Prepend[Append[init, s], s]]; Close[str];
                Run["f77 tstopt1.f -lnag"];
                Run["a.out"];
                results = ReadList["optsol.r", Word, RecordLists -> True];
                res1 = Take[results, {6, (5+nrcn+ncomp)}];
                xstring = Map[Part[#, 4]&, res1];
                strm = StringToStream[StringJoin[ Flatten[Transpose [{xstring, Table["
", {i, ncomp + nrcn}]}] ] ] ];
                ylist = ReadList[strm, Number];
                BarChart[ N[ Exp[ Flatten[Transpose[{init, ylist}]] ] ] ]

        ]


End[ ]
EndPackage[ ]
```

InVivo.m

```
InVivo[aa_, mo_] :=
    Block[{i, alph, xstring},
        ncomp = Length[aa];
        a = Transpose[Drop[Transpose[aa], -(ncomp-mo)]];
        data = ReadList["invivo1.d", Word, RecordLists -> True];
        xstring1 = Map[Part[#, 2]&, Take[data, {4, 10}]];
        xstring2 = Map[Part[#, 1]&, Take[data, {4, 10}]];
        benz = ToExpression[xstring1];
        time = ToExpression[xstring2];
        forprog = ReadList["invivot.f", Character];
        forprog[[164]] = ToString[(Length[aa]+2)];
        forprog[[171]] = ToString[(Length[time]+2)];
        forprog[[821]] = "invivo1.d";
        str1 = OpenWrite["invivo1.f"];
        WriteString[str1, StringJoin[forprog]];Close[str1];
        Run["f77 invivo1.f -lnag"];
        Run["a.out"];
        alph = ReadList["invivo.r", Word, RecordLists -> True];
        xstring = Map[Part[#, 2]&, Take[alph, {7, (6+Length[aa])}]];
        alpha = ToExpression[xstring];
        ji=(benz[[1]]-benz[[7]])/(time[[7]]-time[[1]]);
        j = DiagonalMatrix[Table[ji, {i, 1, Length[Transpose[a]]}]];
        coeffs = alpha.a.j;
        BarChart[coeffs]
    ]
```

fulopt.f

```
*     E04MBF Example Program Text
*     Mark 14 Revised.  NAG Copyright 1989.
*     .. Parameters ..
      INTEGER          NCLIN, NROWA, N, NCTOTL, LIWORK, LWORK
                       P   A   R   A   M   E   T   E   R
(NCLIN=7,NROWA=NCLIN,N=29,NCTOTL=NCLIN+N,
     +
LIWORK=2*N,LWORK=2*N*N+6*N+N+4*NCLIN+NROWA)
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=7,NOUT=8)
*     .. Local Scalars ..
      DOUBLE PRECISION OBJLP
      INTEGER          I, IFAIL, ITMAX, J, MSGLVL
      LOGICAL          LINOBJ
*     .. Local Arrays ..
      DOUBLE PRECISION A(NROWA,N), BL(NCTOTL), BU(NCTOTL),
     +                 CLAMDA(NCTOTL), CVEC(N), WORK(LWORK), X(N)
      INTEGER          ISTATE(NCTOTL), IWORK(LIWORK)
*     .. External Subroutines ..
      EXTERNAL         E04MBF, X04ABF
      OPEN(UNIT = 7, FILE = 'lucy/fulopt.d')
      OPEN(UNIT = 8, FILE = 'lucy/fulopt.r')
*     .. Executable Statements ..
      WRITE (NOUT,*) 'E04MBF Example Program Results'
*     Skip heading in data file
      READ (NIN,*)
      CALL X04ABF(1,NOUT)
      ITMAX = 20
      MSGLVL = 1
```

```
      LINOBJ = .TRUE.
      READ (NIN,*) (CVEC(J),J=1,N)
      READ (NIN,*) ((A(I,J),J=1,N),I=1,NCLIN)
      READ (NIN,*) (BL(J),J=1,NCTOTL)
      READ (NIN,*) (BU(J),J=1,NCTOTL)
      READ (NIN,*) (X(J),J=1,N)
      IFAIL = 1
*
                                                  C     A     L     L
E04MBF(ITMAX,MSGLVL,N,NCLIN,NCTOTL,NROWA,A,BL,BU,CVEC,LINOBJ,
                                                 +
X,ISTATE,OBJLP,CLAMDA,IWORK,LIWORK,WORK,LWORK,IFAIL)
*
      STOP
      END
```

MetData.m

```
BeginPackage["InputData3`"]

InputData3::usage = "contains data"


aa = { {-1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,-1, 0, 0, 0},
       { 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,-1, 0,-1, 0, 0},
       { 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,-1, 0, 0,-1, 0},
       { 0, 0, 0, 0, 0, 0, 0, 0, 1,-1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,-1},
       { 1,-1, 0, 0, 0, 0, 0, 0, 0, 0,-1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0},
       { 0, 1,-1, 0, 0, 0, 0, 0, 0, 0, 0,-1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0},
       { 0, 0, 1,-1, 0, 0, 0, 0,-1, 1, 0, 0,-1, 1, 0, 0, 0, 0, 0, 0, 0, 0},
       { 0, 0, 0, 1,-1, 0, 0, 0, 0, 0, 0, 0, 0,-1, 1, 0, 0, 0, 0, 0, 0, 0},
       { 0, 0, 0, 0, 1,-1, 0, 0, 0, 0, 0, 0, 0, 0,-1, 1, 0, 0, 0, 0, 0, 0},
       { 0, 0, 0, 0, 0, 1,-1, 0, 1,-1, 0, 0, 0, 0, 0,-1, 1, 0, 0, 0, 0, 0},
       { 0, 0, 0, 0, 0, 0, 1,-1, 0, 0, 0, 0, 0, 0, 0, 0,-1, 1, 0, 0, 0, 0}    };

k = {       {      7.62,  75                              },
      {      2,     0.20                          },
      {      50,    3.0                           },
      {      100,   2.0                           },
      {      6.00,  20                            },
      {      2.6,   15                            },
      {      5.2,   30                            },
      {      7.8,   53                            },
      {      3.2,   100                           },
      {      2.6,   74,     31,     49,    62     },
      {      1,     18                            },
      {      0.26,  15                            },
      {      0.03,  23                            },
```

```
        {       0.52,  30                              },
        {       0.11,  21                              },
        {       0.78,  53                              },
        {       0.32,  100                             },
        {       0.26,  74,     31,     49,     62      }       };
int = {{x[5], 20.}, {x[6], 25.}, {x[7], 33.}, {x[8], 41.}, {x[9], 32.}, {x[10], 29},
{x[11], 30}};
ext = {{x[1], 30., 20., 40.}, {x[2], 15.}, {x[3], 19.}, {x[4], 30.}};
enzime = {{e[1], 210.}, {e[2], 210.}, {e[3], 200.}, {e[4], 195.}, {e[5], 180.}, {e[6],
185.}, {e[7], 210.}, {e[8], 210.}, {e[9], 100.}, {e[10], 100.}, {e[11], 195.}, {e[12],
195.}, {e[13], 230.}, {e[14], 230.}, {e[15], 128.}, {e[16], 128.}, {e[17], 250.}, {e[18],
250.} };
EndPackage[]
```

invivo1.d

```
Transient Conc. Data for InVivo
 7 4 'U' 'M'
        Time   X1     X2     X3     X4
        0      10     0      0      0
        5      8      0.1    0.1    0.08
        10     6      0.2    0.2    0.5
        15     5      0.4    0.4    1.0
        20     4      0.5    0.5    1.4
        25     3.5    0.7    0.7    1.6
        30     2.8    1.0    1.0    2.0
        1      1      1      1
```

roots.f

```
*     C05NBF Example Program Text
*     Mark 14 Revised.  NAG Copyright 1989.
*     .. Parameters ..
      INTEGER          N, LWA
      PARAMETER        (N=1,LWA=(N*(3*N+13))/2)
      INTEGER          NOUT
      PARAMETER        (NOUT=8)
*     .. Local Scalars ..
      DOUBLE PRECISION FNORM, TOL
      INTEGER          I, IFAIL, J
*     .. Local Arrays ..
      DOUBLE PRECISION FVEC(N), WA(LWA), X(N)
*     .. External Functions ..
      DOUBLE PRECISION F06EJF, X02AJF
      EXTERNAL         F06EJF, X02AJF
*     .. External Subroutines ..
      EXTERNAL         C05NBF, FCN
*     .. Intrinsic Functions ..
      INTRINSIC        SQRT
      OPEN(UNIT = 8, FILE = 'lucy/roots.r')
*     .. Executable Statements ..
      WRITE (NOUT,*) 'C05NBF Example Program Results'
      WRITE (NOUT,*)
*     The following starting values provide a rough solution.
      DO 20 J = 1, N
        X(J) = 1.0D0
   20 CONTINUE
      TOL = SQRT(X02AJF())
      IFAIL = 1
*
      CALL C05NBF(FCN,N,X,FVEC,TOL,WA,LWA,IFAIL)
```

```
*
      IF (IFAIL.EQ.0) THEN
        FNORM = F06EJF(N,FVEC,1)
        WRITE (NOUT,99999) 'Final 2-norm of the residuals =', FNORM
        WRITE (NOUT,*)
        WRITE (NOUT,*) 'Final approximate solution'
        WRITE (NOUT,*)
        WRITE (NOUT,99998) (X(J),J=1,N)
      ELSE
        WRITE (NOUT,99997) 'IFAIL = ', IFAIL
        IF (IFAIL.GT.1) THEN
          WRITE (NOUT,*)
          WRITE (NOUT,*) 'Approximate solution'
          WRITE (NOUT,*)
          WRITE (NOUT,99998) (X(I),I=1,N)
        END IF
      END IF
      STOP
*
99999 FORMAT (1X,A,D12.4)
99998 FORMAT (1X,3F12.4)
99997 FORMAT (1X,A,I2)
      END
*
      SUBROUTINE FCN(N,X,FVEC,IFLAG)
*     .. Scalar Arguments ..
      INTEGER          IFLAG, N
*     .. Array Arguments ..
      DOUBLE PRECISION FVEC(N), X(N)
*     .. Local Scalars ..
*     .. Executable Statements ..
      FVEC(1) = 801.967 - 1000*X(4)/(3 + X(1)) - 1400.*X(4)/(8 + X(1))
```

```
      RETURN
      END
```

tstopt.f

```
*     E04MBF Example Program Text
*     Mark 14 Revised.  NAG Copyright 1989.
*     .. Parameters ..
      INTEGER          NCLIN, NROWA, N, NCTOTL, LIWORK, LWORK, NALL
      PARAMETER        (NCLIN= ,NROWA=NCLIN,N= ,NCTOTL=NCLIN+N,
     +
LIWORK=2*N,LWORK=2*N*N+6*N+N+4*NCLIN+NROWA,
     +                 NALL=NROWA*N)
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=7,NOUT=8)
*     .. Local Scalars ..
      DOUBLE PRECISION OBJLP
      INTEGER          I, IFAIL, ITMAX, J, MSGLVL
      LOGICAL          LINOBJ
*     .. Local Arrays ..
      DOUBLE PRECISION A(NROWA,N), AA(NALL), BL(NCTOTL),
BU(NCTOTL),
     +                 CLAMDA(NCTOTL), CVEC(N), WORK(LWORK), X(N)
      INTEGER          ISTATE(NCTOTL), IWORK(LIWORK)
      CHARACTER A1, A2
*     .. External Subroutines ..
      EXTERNAL         E04MBF, X04ABF
      OPEN(UNIT = 7, FILE = 'modsol.r')
      OPEN(UNIT = 8, FILE = 'optsol.r')
*     .. Executable Statements ..
      WRITE (NOUT,*) 'E04MBF Example Program Results'
*     Skip heading in data file
```

```
*     READ (NIN,*)
      CALL X04ABF(1,NOUT)
      ITMAX = 20
      MSGLVL = 1
      LINOBJ = .TRUE.
      READ (NIN,*) A1, (CVEC(J),J=1,N), A2
      READ (NIN,*) A1, (AA(J),J=1,NALL), A2
      READ (NIN,*) A1, (BL(J),J=1,NCTOTL), A2
      READ (NIN,*) A1, (BU(J),J=1,NCTOTL), A2
      READ (NIN,*) A1, (X(J),J=1,N), A2
      DO 20 I = 1,NCLIN
        DO 30 J = 1,N
              A(I,J) = AA(J + 29*(I-1))
30      CONTINUE
20    CONTINUE
      IFAIL = 1
*
                                                C       A       L       L
E04MBF(ITMAX,MSGLVL,N,NCLIN,NCTOTL,NROWA,A,BL,BU,CVEC,LINOBJ,
                                                +
X,ISTATE,OBJLP,CLAMDA,IWORK,LIWORK,WORK,LWORK,IFAIL)
*
      STOP
      END
```

Appendix E: Carbon Balance for Calculation of By-Product Formation

By measuring the primary carbon-containing components in the fermentation system and performing a balance over them, it is possible to calculate how much of the carbon feed is being channelled into unwanted by-products. By using total carbon measurements, the problem of analysing for specific by-products is avoided, although is necessary if the balance is to be refined and if further physiological information is required.

The carbon balance is based on two conservation relationships:

$$\text{Total Carbon}_{broth} = C_{benzoate} + C_{glycerol} + C_{biomass} + C_{byproducts}$$

$$\Delta C_{benzoate} + \Delta C_{glycerol} = \Delta C_{biomass} + \Delta C_{CO2} + \Delta C_{byproducts}$$

From these two relationships, the two unknown quantities, $C_{glycerol}$ and $C_{byproducts}$ can be calculated using the expressions:

$$\Delta C_{byproducts} = 1/2\ \Delta_{TCbroth} - \Delta C_{biomass} - 1/2\ \Delta C_{CO2}$$

$$\Delta C_{glycerol} = \Delta C_{biomass} + \Delta C_{CO2} + \Delta C_{byproducts} - \Delta C_{benzoate}$$

In this case, glycerol concentration is easy to assay by other means (see 3.4.4). However if the microorganism was growing on complex media this method would give the uptake of carbon source. Also, it is only by a carbon balance such as this that the carbon lost to byproducts can be calculated.

For cells operating under fully aerobic conditions where the TCA cycle is functional, the following relationship should also hold on a molar basis:

$$\Delta\text{glycerol} = 1/3\ \Delta CO_2$$

These relationships are examined below using the total carbon data described in Section 4.4. A LabView interface, **Carbon Fluxes** has been written to calculate the results of these balances (see Appendix E for documentation). This program calculates the carbon fluxes and the specific growth of the organism and these are then displayed in strip chart form if the data analysis is being performed on-line or on a graph for the analysis of historic data.

Results

Using the above balance method, the build-up of byproducts was estimated for

fermentations of *E.coli* pQR150, *E.coli* pQR185 and *E.coli* pQR185 under the different fermentation conditions described in Section 3 and whose fermentation performance was discussed in Section 4.3. The results obtained for *E.coli* pQR150 are shown in Figure E.1. In this figure it can be seen that the level of by-product formation is greatest when the strain is grown on M9 without plasmid induction or presence of the substrate benzoate, and least when the pathway is induced by IPTG and supplied with benzoate. From these results the correlation between growth rate and by-product formation is very high. The outliers in the Total Carbon data as discussed in Section 4.5 are reflected here in the profile for growth on M9/IPTG and to a lesser extent in that for growth on M9. The accumulation of by-products over the stationary phase is evident in the profiles for growth on M9/benzoate and on M9/benzoate/IPTG.

The profiles for by-product formation during growth of E.coli pQR185, shown in Figure E.2 are slightly different. While roughly the same final levels are reached, they do not show the same correspondence with growth rate. For example, growth of the strain on M9/benzoate/ IPTG results in higher by-product levels as compared to those achieved during growth on M9 alone even though the final cell density achieved on M9 was higher than that for M9/benzoate/IPTG (this can be seen in Figure 4.4). However the initial growth rate on M9/benzoate/IPTG is higher than that on M9 only so it is probable that this is the cause for the by-product build-up. This indicates that, in this case, it is the initial growth rate and not the final cell density which determines the by-product levels and that by-product formation occurs primarily during exponential growth.

Examining the profiles of by-product levels during growth of *E.coli* pQR186 in Figure E.3, it can be seen that they correspond very closely to the growth profiles in Figure 4.7, in that by-product levels are initially higher during growth on M9/benzoate/IPTG when the growth is higher but these are exceeded by the levels obtained during growth on M9 after 14 hours. The highest by-product levels obtained for *E.coli* pQR186 were roughly half that of those for *E.coli* pQR150 and *E.coli* pQR185.

The purpose of the glycerol estimator in Eqn.85 is to provide a rapid estimate of the glycerol levels in the fermentation without the need for an additional assay. Because it will contain all of the errors in the individual variables in the right hand side of Eqn.85 it will be less accurate than the enzyme assay described in Section 3.6. The actual and

estimated glycerol profiles for growth of *E.coli* pQR150 on M9 are shown in Figure E.4. It can be seen that the estimate is initially good but deteriorates over the time course of the fermentation as the errors accumulate. One probable source of error which has been previously mentioned is a poor calibration of the mass spectrometer pumps. This is illustrated in the plot of Figure 4.13 where the slope of the line is -1.29 where it ought to be -1.0. If this is corrected for then the estimate improves to that shown in Figure E.5. Also shown in Figure E.5 is the glycerol estimate obtained from Eqn.86 above.

Errors in this approximation can occur for several reasons: one is the previously mentioned inaccuracies in the off-gas analysis, and another is the probability of glycerol or intermediates being taken off via other pathways.

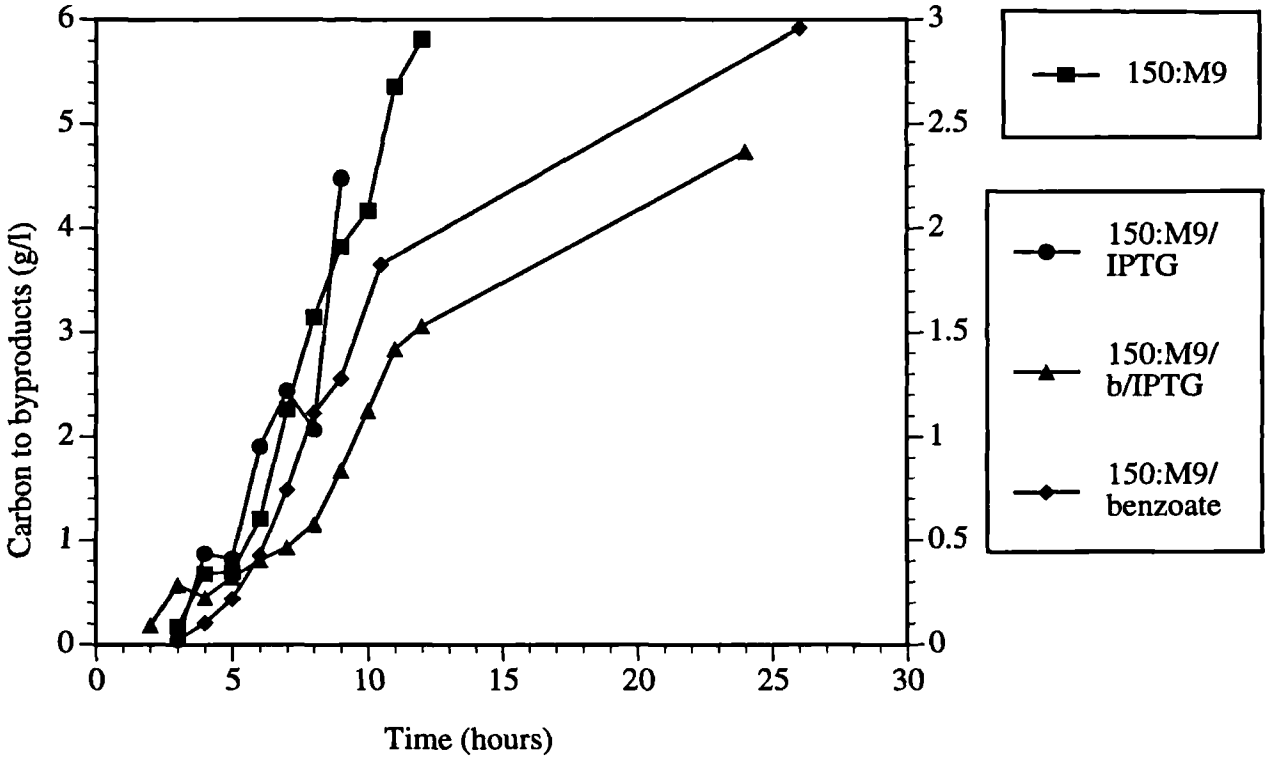


Figure E.1: Accumulation of by-products during fermentation of *E.coli* pQR150 under different fermentation conditions: M9 minimal media, M9 with IPTG for plasmid induction, M9 with IPTG plus the pathway substrate benzoate (b) and M9 with benzoate but without plasmid induction.

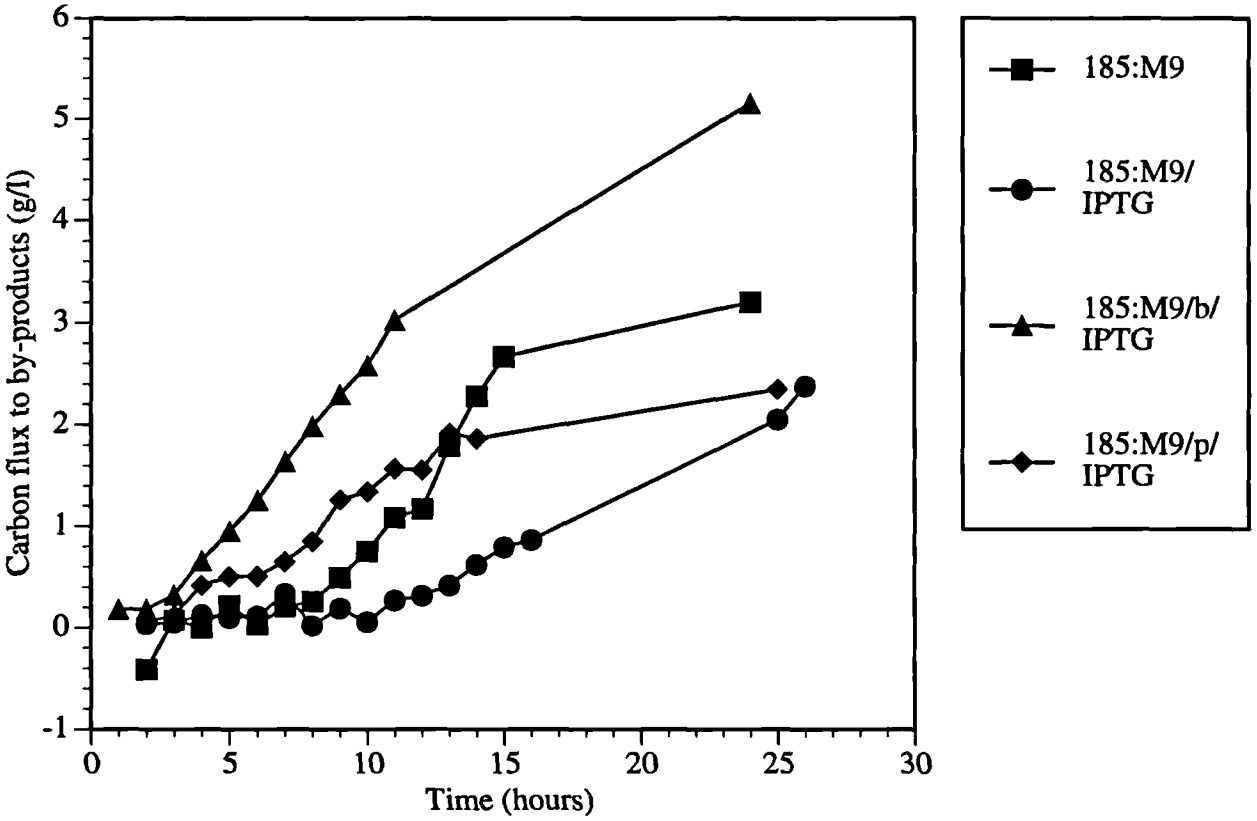


Figure E.2: Accumulation of by-products during fermentation of *E.coli* pQR185 under different fermentation conditions: M9 minimal media, M9 with IPTG for plasmid induction, M9 with IPTG plus the pathway substrate benzoate (b) and M9 with IPTG plus the pathway substrate p-toluate (p).

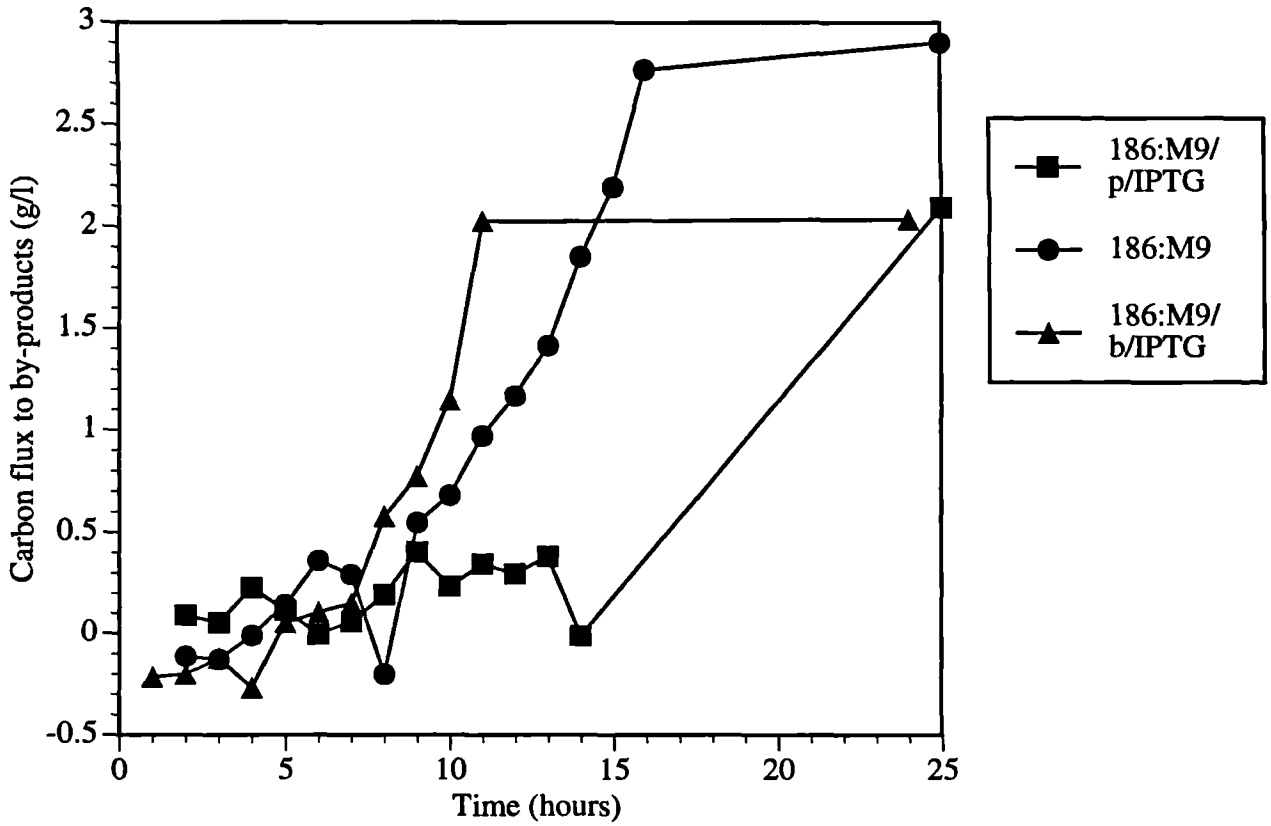


Figure E.3: Accumulation of by-products during fermentation of *E.coli* pQR186 under different fermentation conditions: M9 minimal media, M9 with IPTG plus the pathway substrate benzoate (b) and M9 with IPTG plus the pathway substrate p-toluate (p).

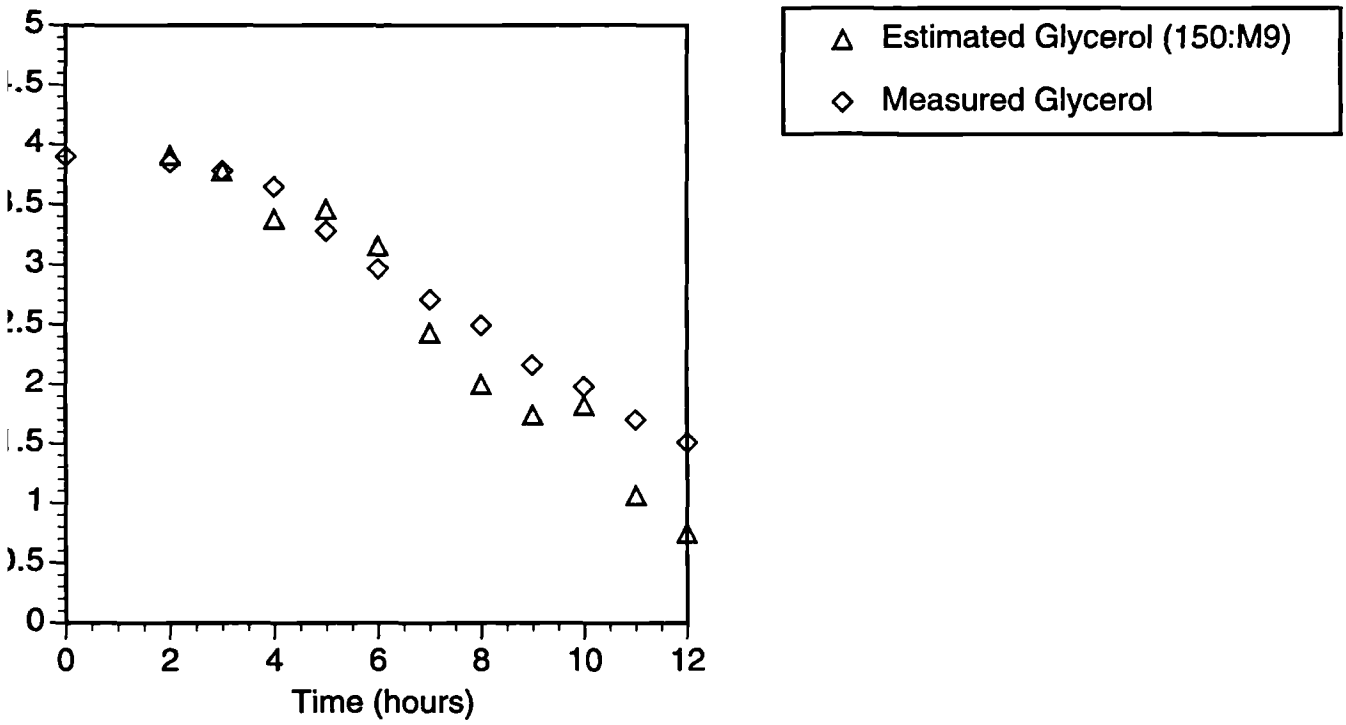


Figure E.4: Estimate of carbon source from total carbon data.

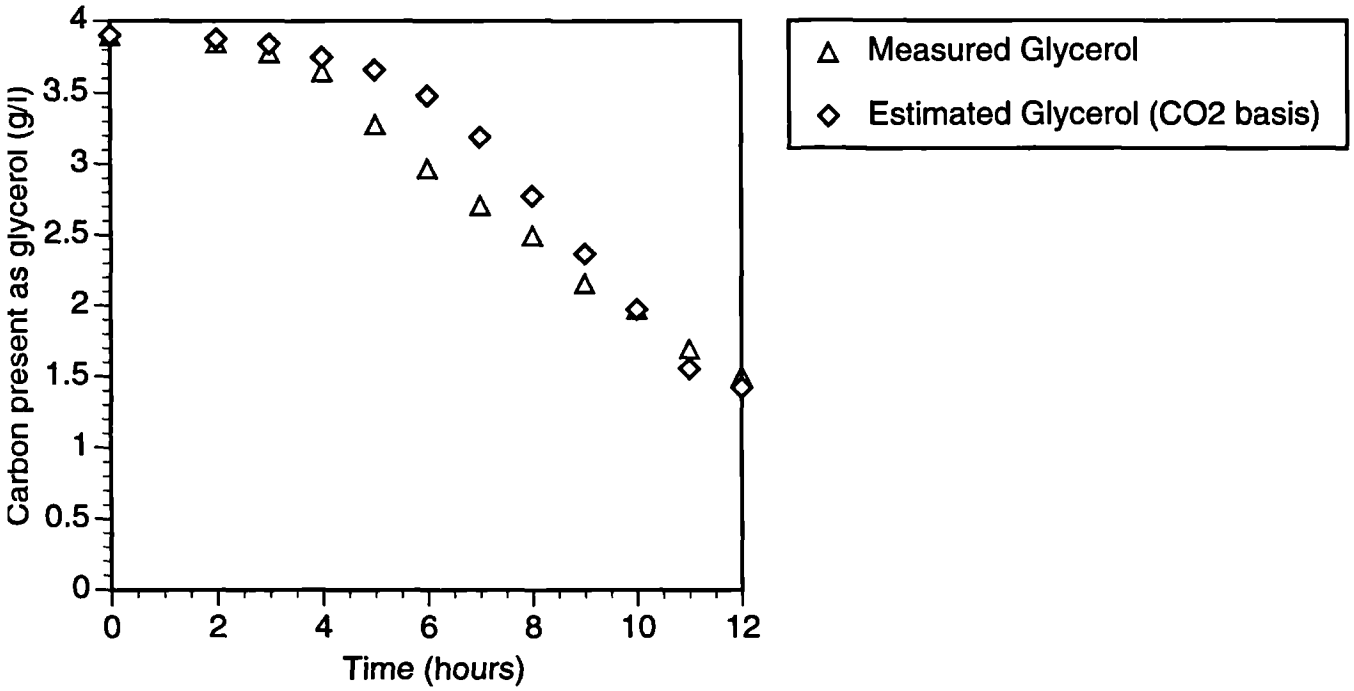
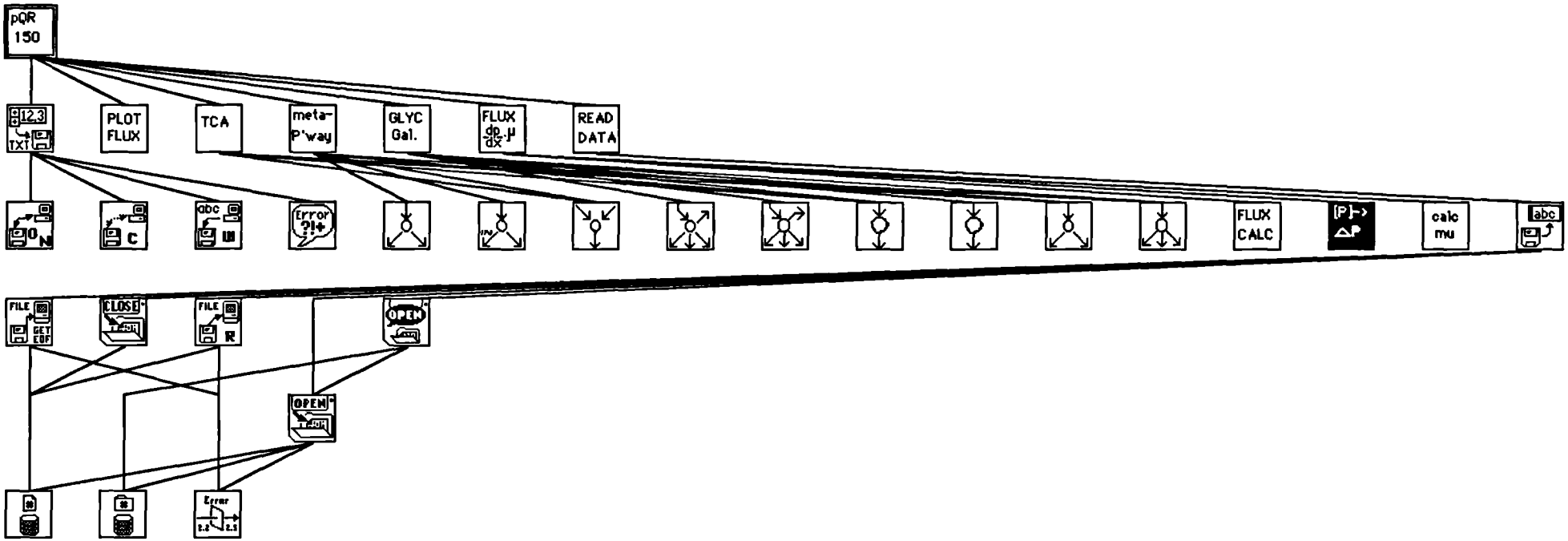


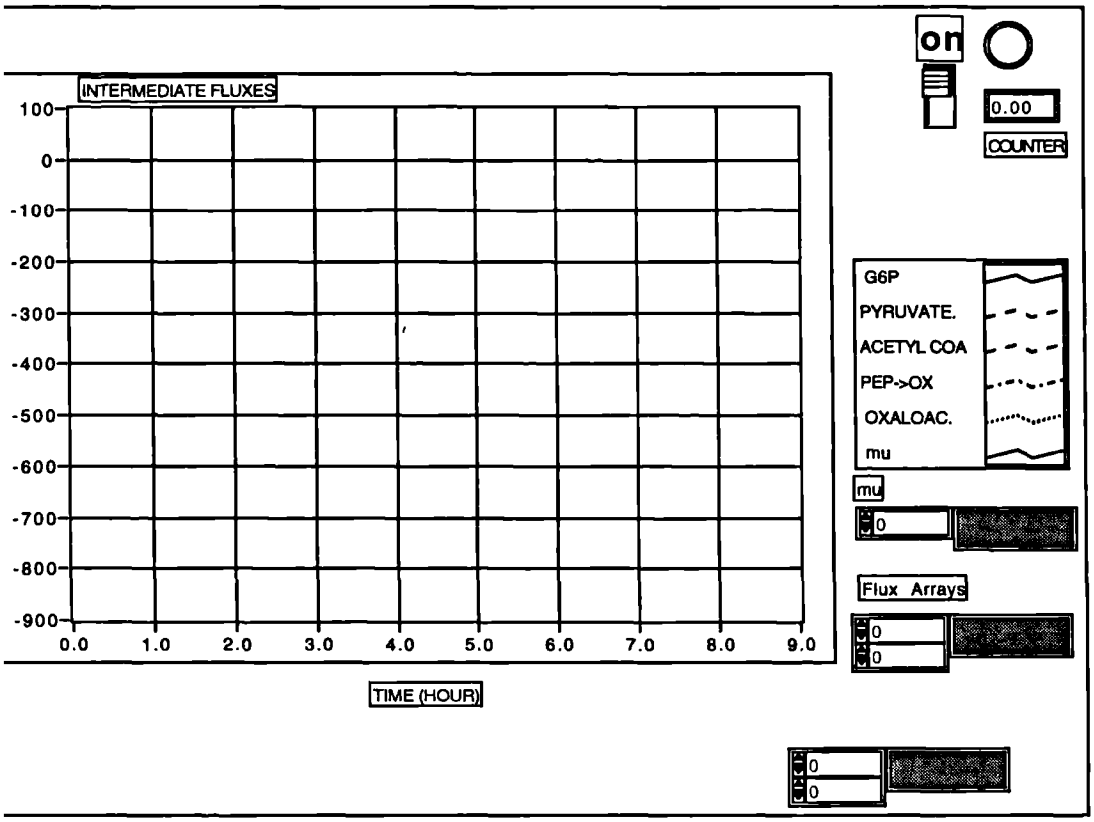
Figure E.5: Estimate of carbon source from CO2 data.

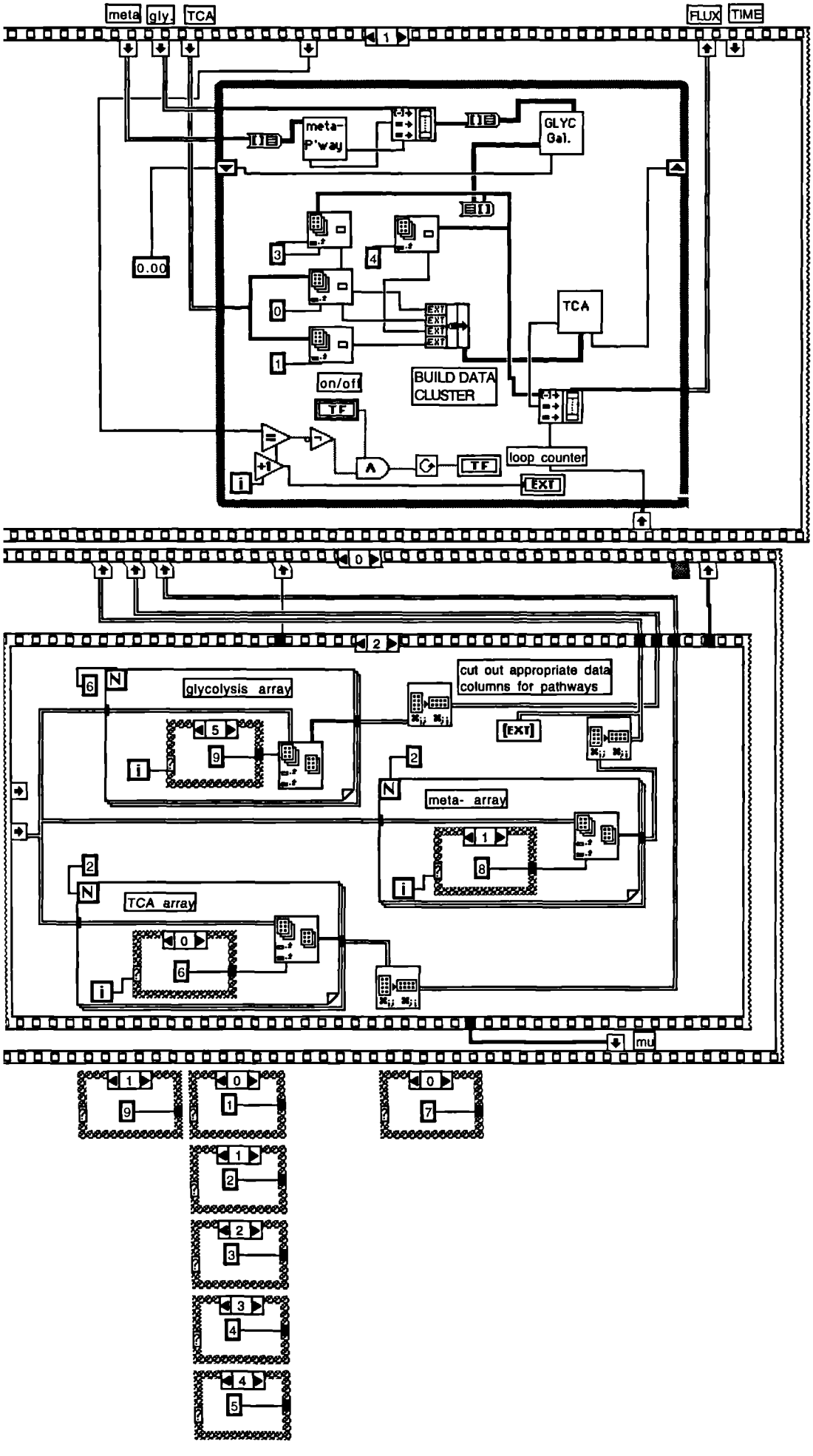
Position in Hierarchy

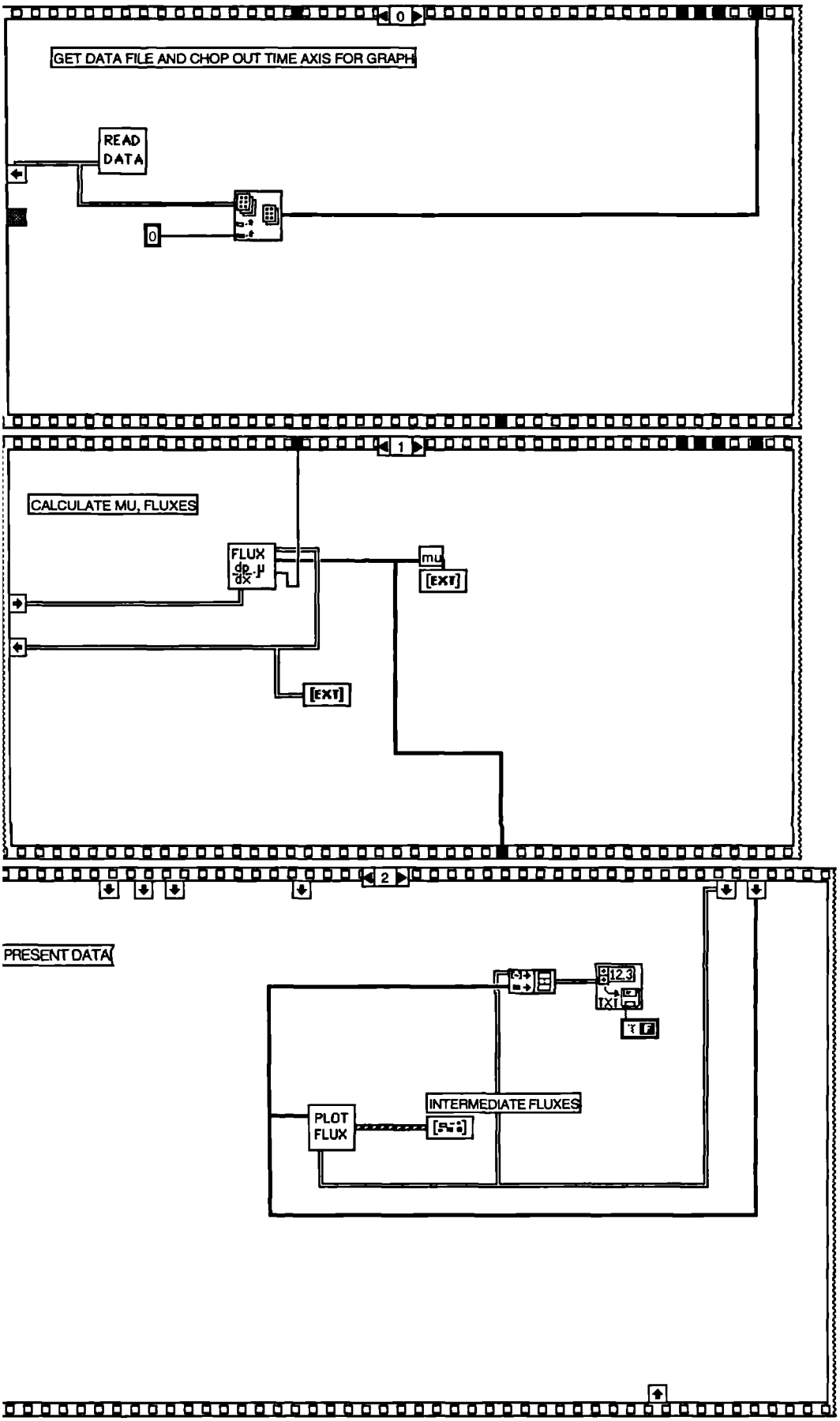


Connector Pane

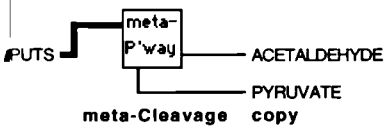
pQR150


INTERMEDIATE FLUXES
100
0
-100
-200
-300
-400
-500
-600
-700
-800
-900
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
TIME (HOUR)
on
0.00
COUNTER
G6P
PYRUVATE.
ACETYL COA
PEP->OX
OXALOAC.
mu
mu
0
Flux Arrays
0
0
0
0

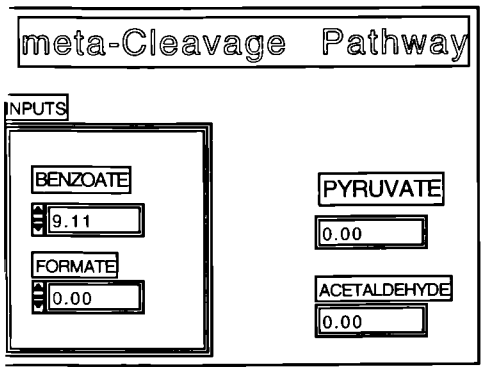

meta
gly
TCA
FLUX
TIME
meta-
P'way
GLYC
Gal.
0.00
TCA
on/off
BUILD DATA
CLUSTER
TF
loop counter
EXT
glycolysis array
cut out appropriate data
columns for pathways
meta- array
TCA array
mu


0
GET DATA FILE AND CHOP OUT TIME AXIS FOR GRAPH
READ DATA
0
1
CALCULATE MU, FLUXES
FLUX
dp/dx
mu
[EXT]
[EXT]
2
PRESENT DATA
123
TXT
INTERMEDIATE FLUXES
PLOT FLUX

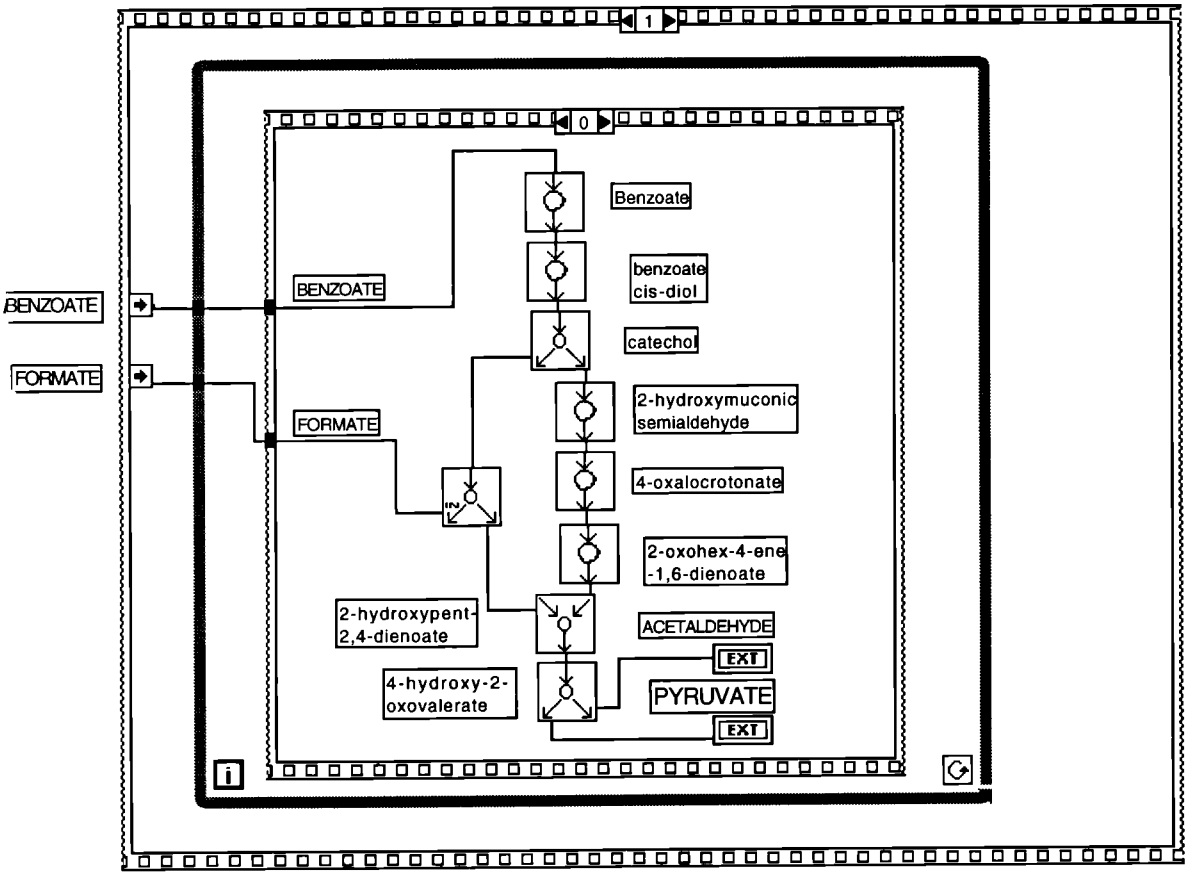
tor Pane

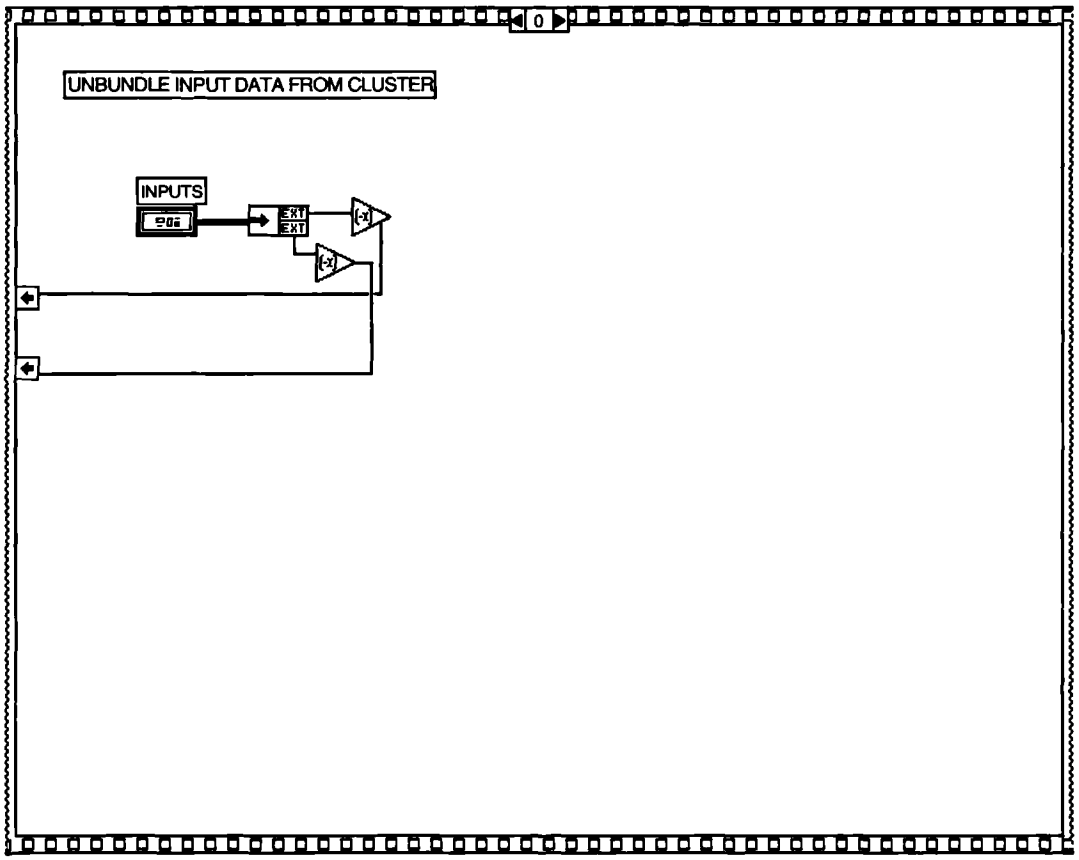


Panel



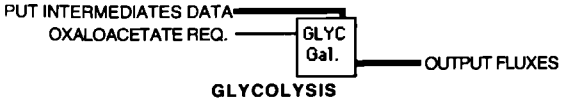
Diagram



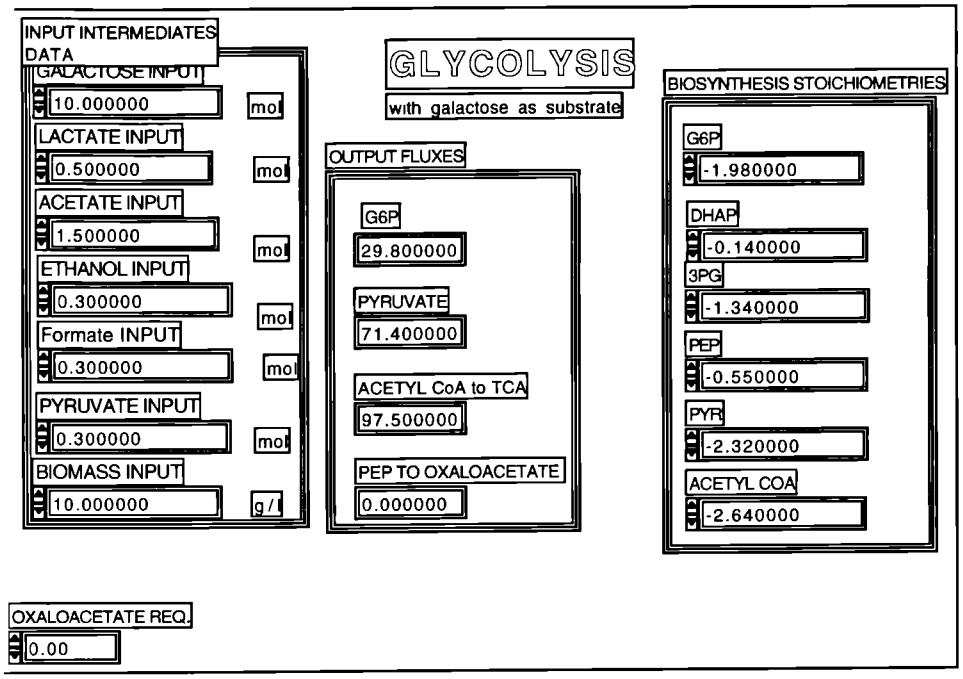

0
UNBUNDLE INPUT DATA FROM CLUSTER
INPUTS

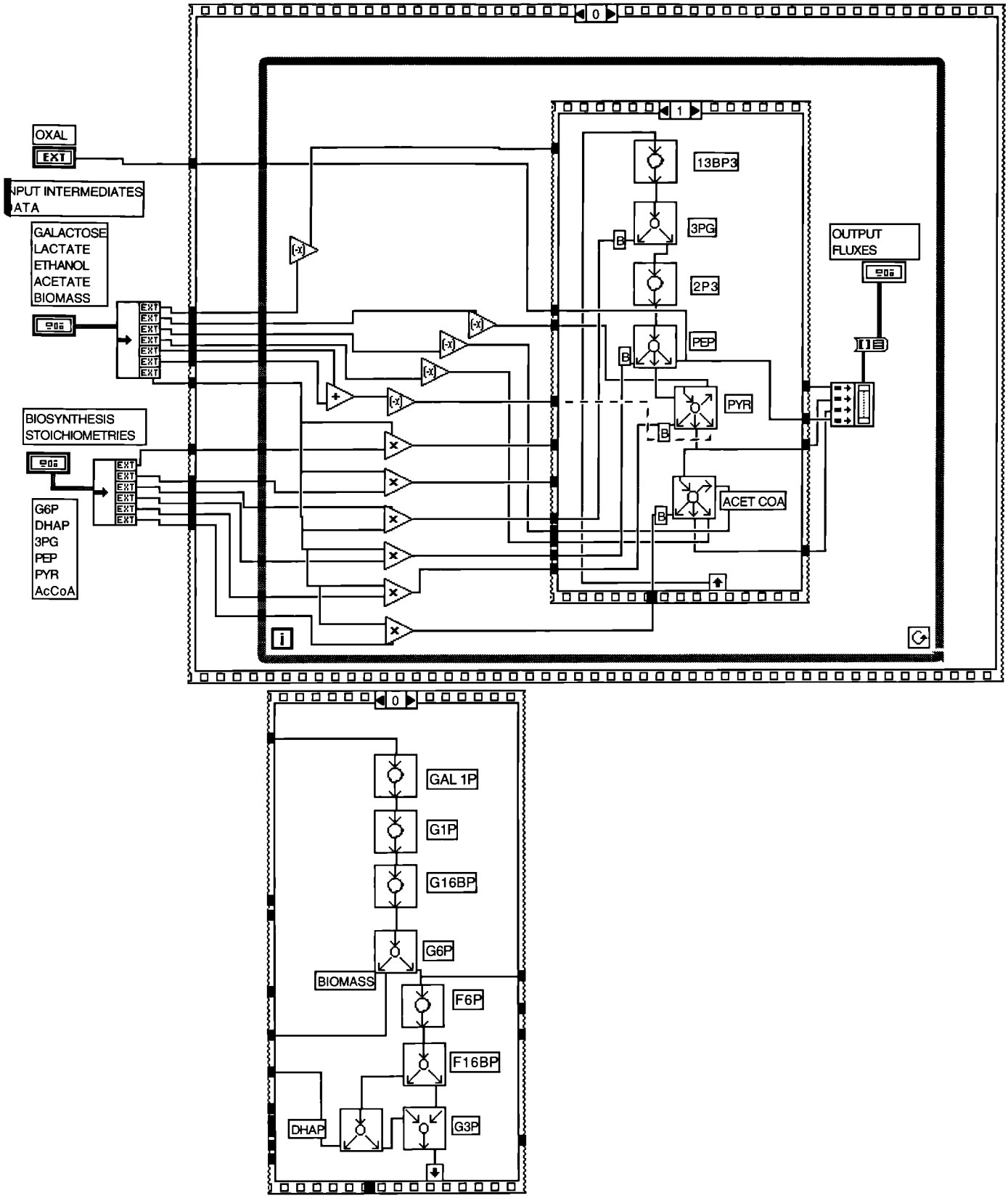
.YSIS
20:40

:tor Pane



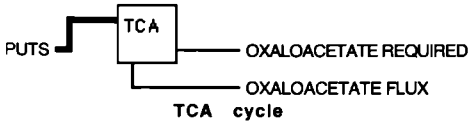
'anel



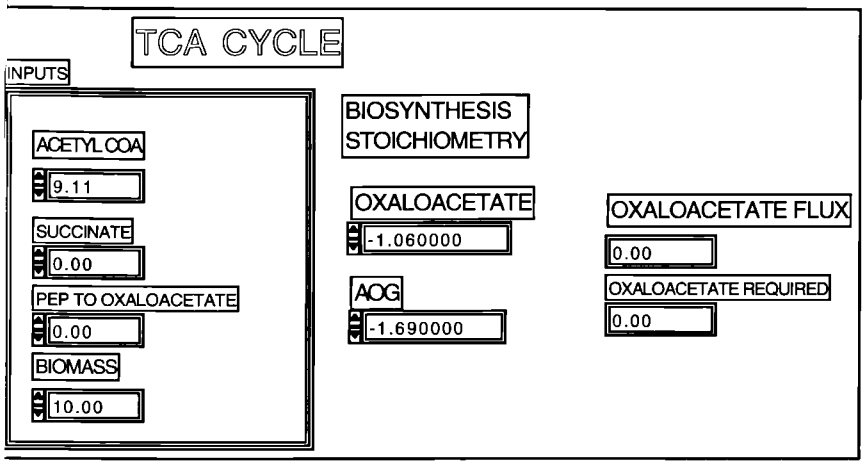

0
1
OXAL
EXT
NPUT INTERMEDIATES
ATA
GALACTOSE
LACTATE
ETHANOL
ACETATE
BIOMASS
BIOSYNTHESIS
STOICHIOMETRIES
G6P
DHAP
3PG
PEP
PYR
AcCoA
13BP3
3PG
2P3
PEP
PYR
ACET COA
OUTPUT
FLUXES
0
GAL 1P
G1P
G16BP
G6P
BIOMASS
F6P
F16BP
DHAP
G3P

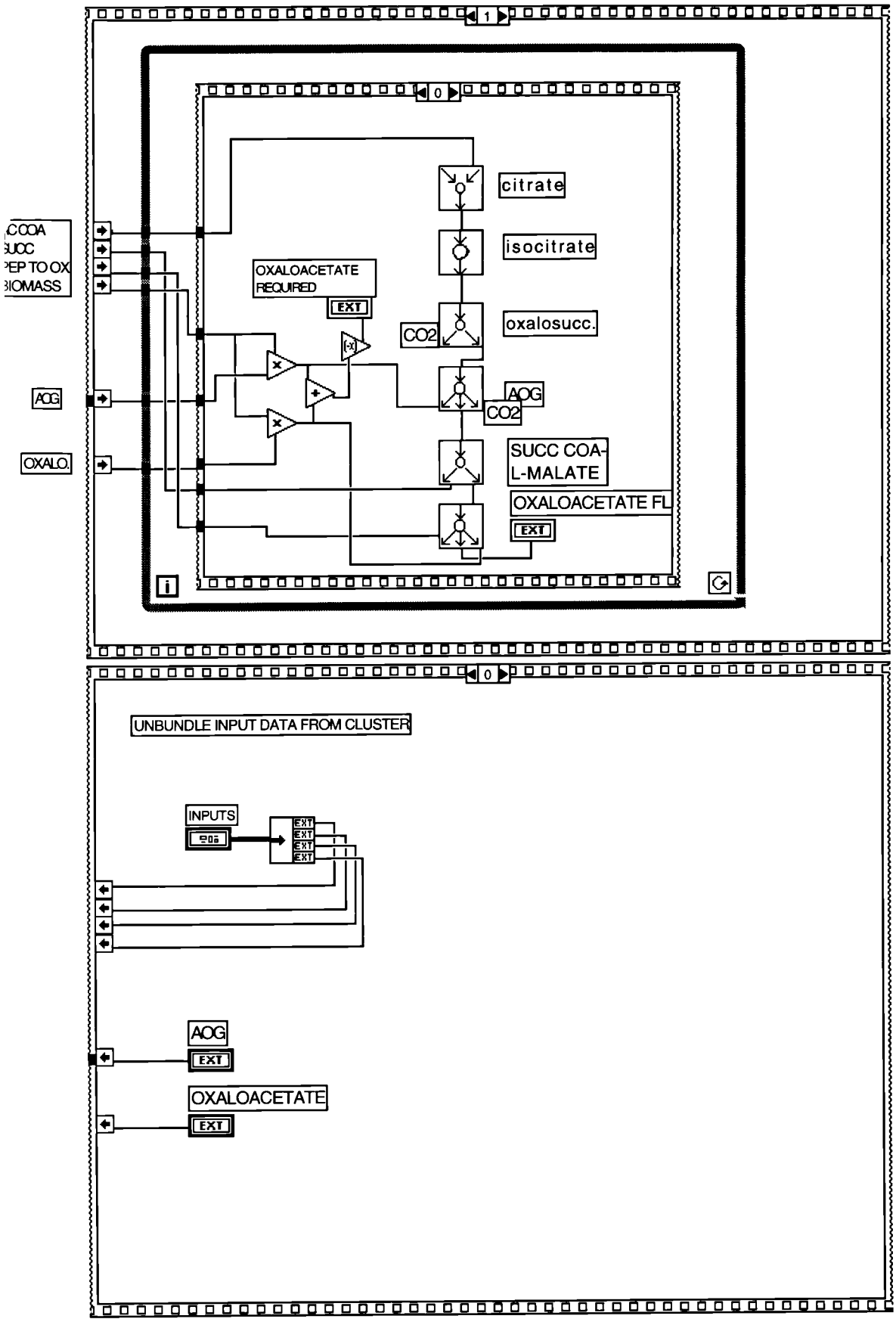
ycle
4 18:11

:tor Pane



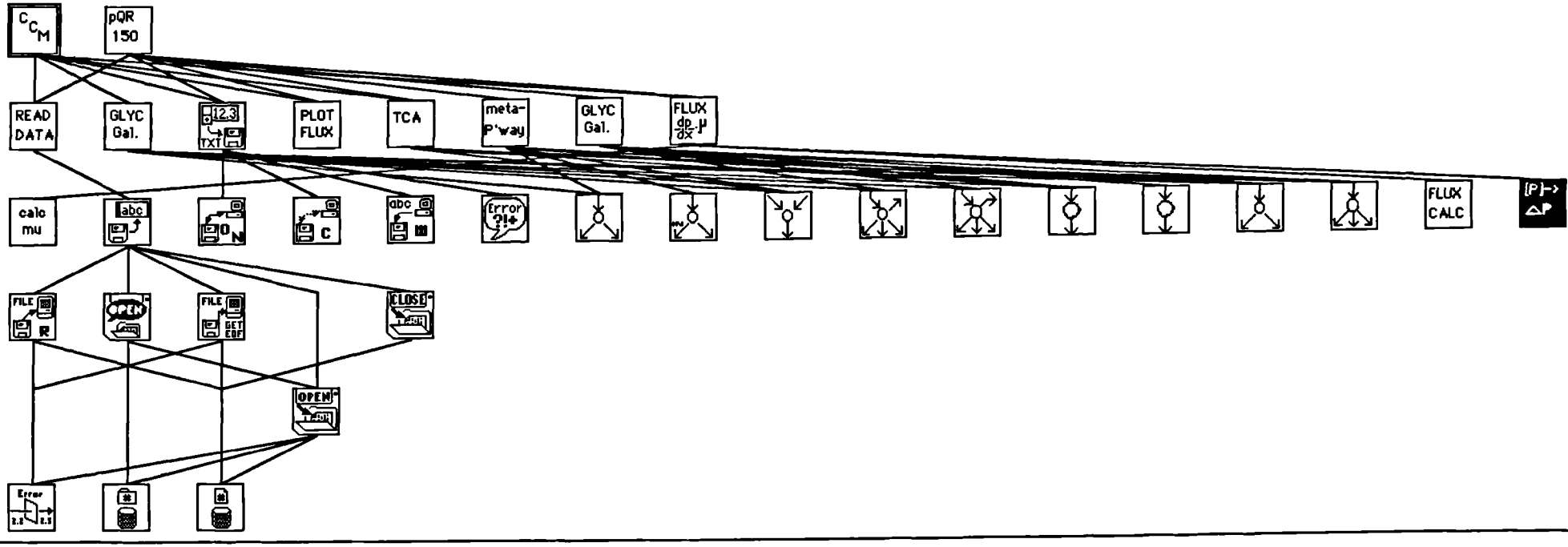
Panel




1
0
citrate
isocitrate
OXALOACETATE REQUIRED
EXT
CO2
oxalosucc.
AOG
CO2
SUCC COA-
L-MALATE
OXALOACETATE FL
EXT
CCOA
UCC
EP TO OX
IOMASS
AOG
OXALO
0
UNBUNDLE INPUT DATA FROM CLUSTER
INPUTS
EXT
EXT
EXT
EXT
AOG
EXT
OXALOACETATE
EXT

CENTRAL CARBON
5/7/94 0:24

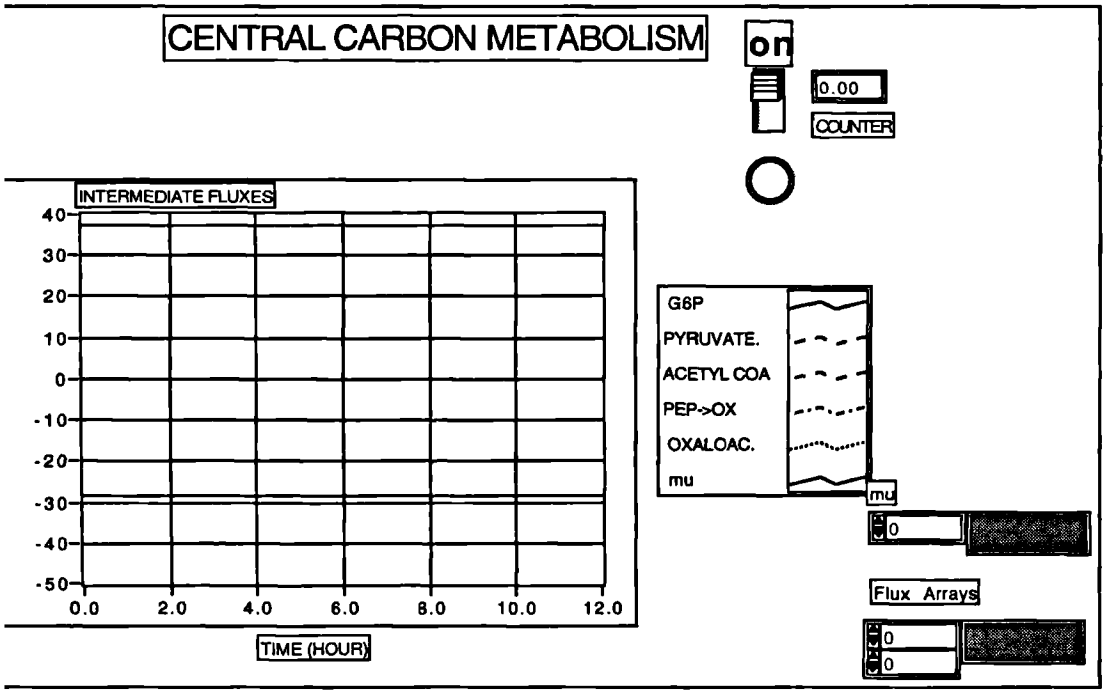
Position in Hierarchy



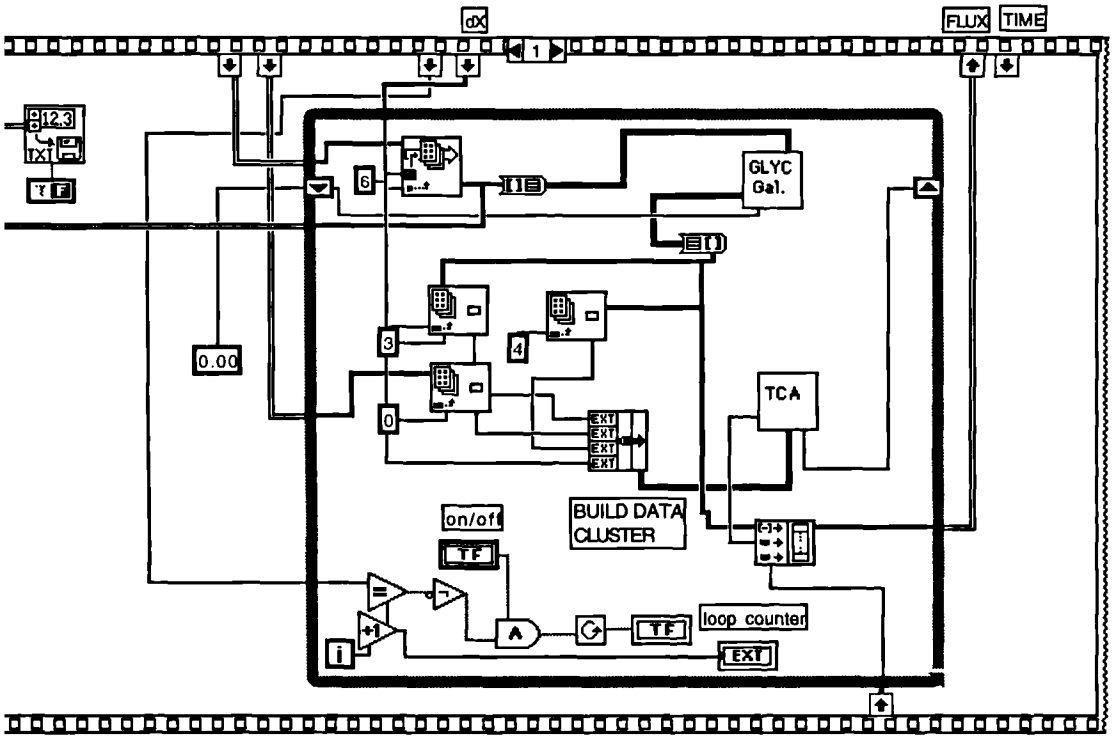
Connector Pane

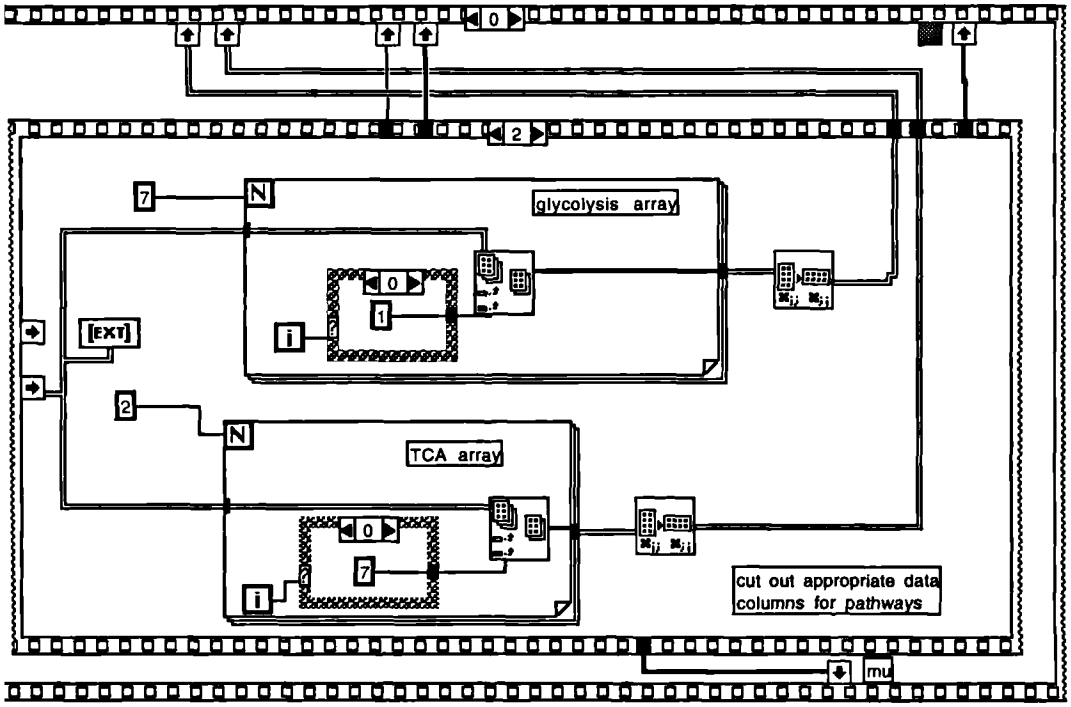
CENTRAL CARBON

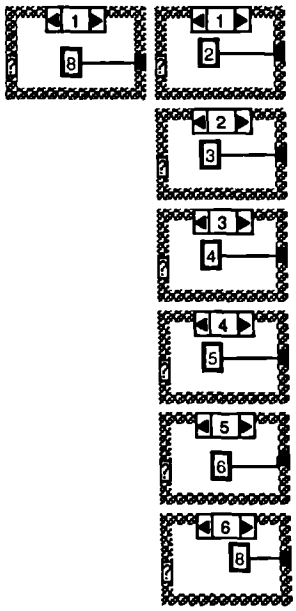
This VI calculates the fluxes in the cetral carbon pathway (glycolysis and the TCA cycle), during growth of E.coli on glucose or galactose. The data required are the input and output concentrations over the time course of the fermentation.

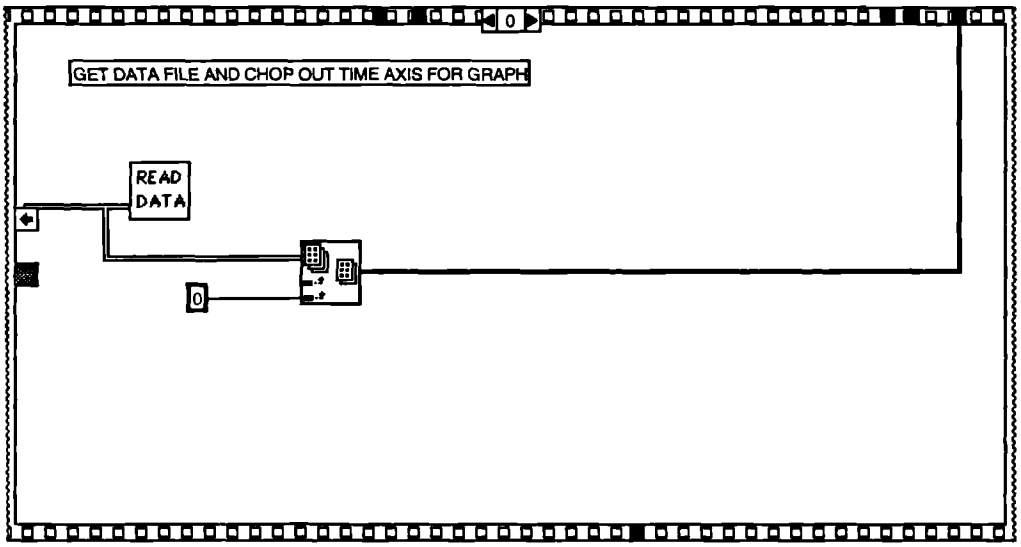

CENTRAL CARBON METABOLISM
on
0.00
COUNTER
INTERMEDIATE FLUXES
40
30
20
10
0
-10
-20
-30
-40
-50
0.0
2.0
4.0
6.0
8.0
10.0
12.0
TIME (HOUR)
G6P
PYRUVATE.
ACETYL COA
PEP->OX
OXALOAC.
mu
mu
0
Flux Arrays
0
0

gram

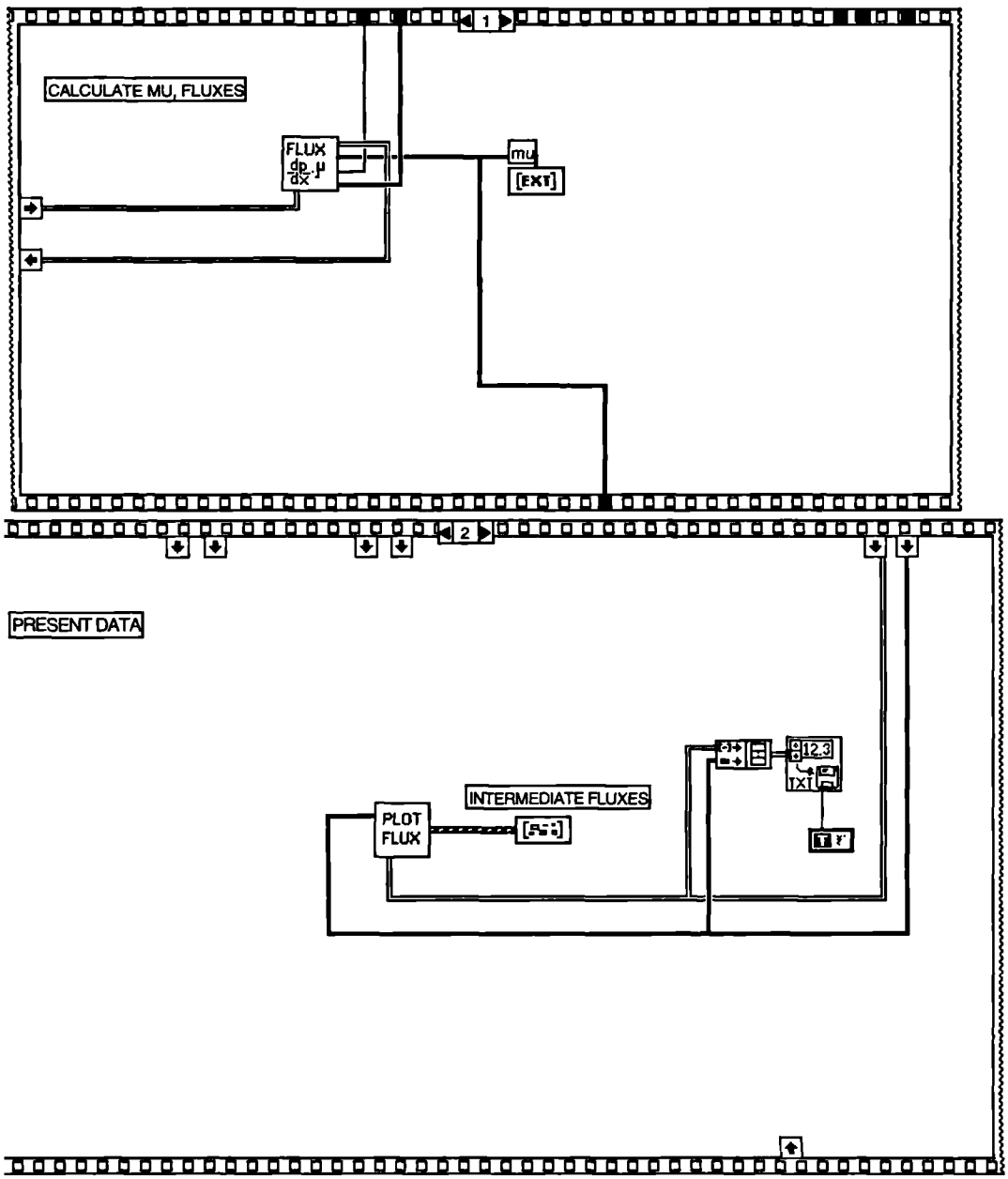

dX
FLUX
TIME
1
12.3
TXT
TF
6
GLYC
Gal.
3
4
0.00
0
TCA
EXT
EXT
EXT
EXT
on/off
TF
BUILD DATA
CLUSTER
=
-1
+1
A
TF
loop counter
EXT
i


0
2
7
N
glycolysis array
0
i
i
[EXT]
2
N
TCA array
0
7
i
cut out appropriate data
columns for pathways
mu


1
8
1
2
2
3
3
4
4
5
5
6
6
8

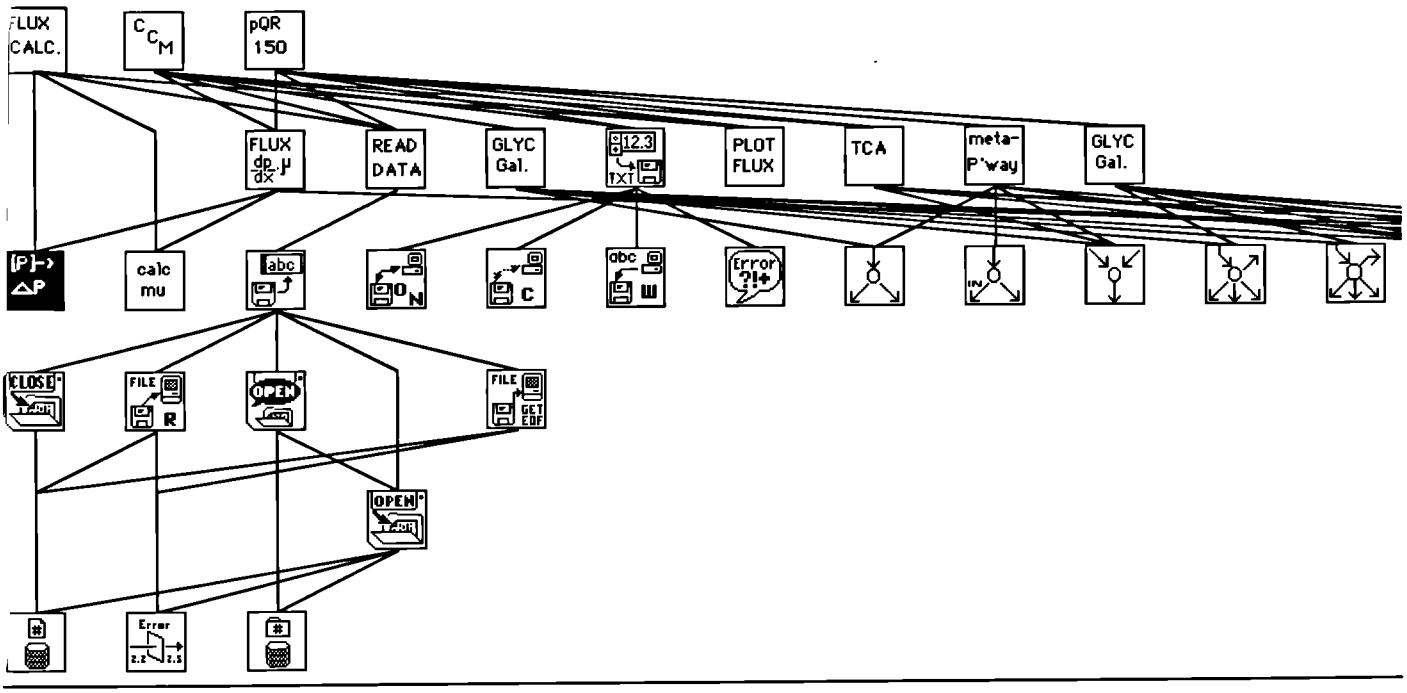

0
GET DATA FILE AND CHOP OUT TIME AXIS FOR GRAPH
READ
DATA
0

CARBON
0:24


1
CALCULATE MU, FLUXES
FLUX
mu
[EXT]
2
PRESENT DATA
INTERMEDIATE FLUXES
PLOT FLUX
12.3
TXT

n Fluxes
23:19

on in Hierarchy

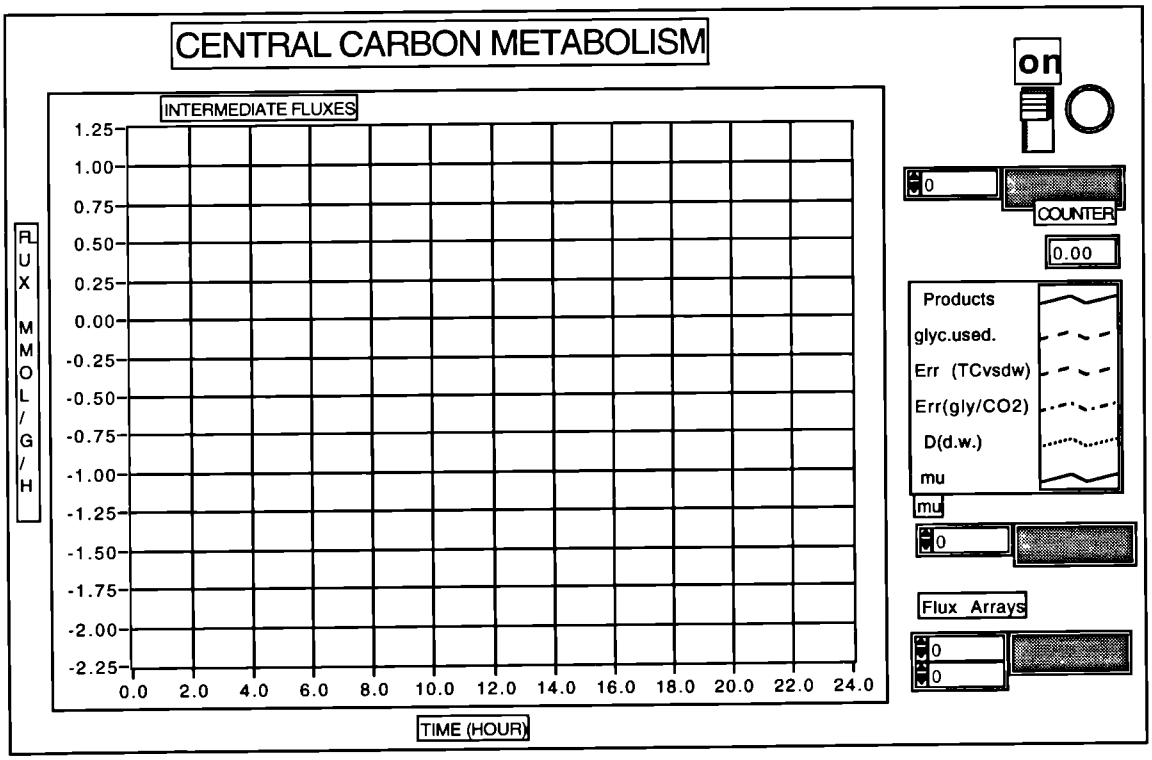


ector Pane

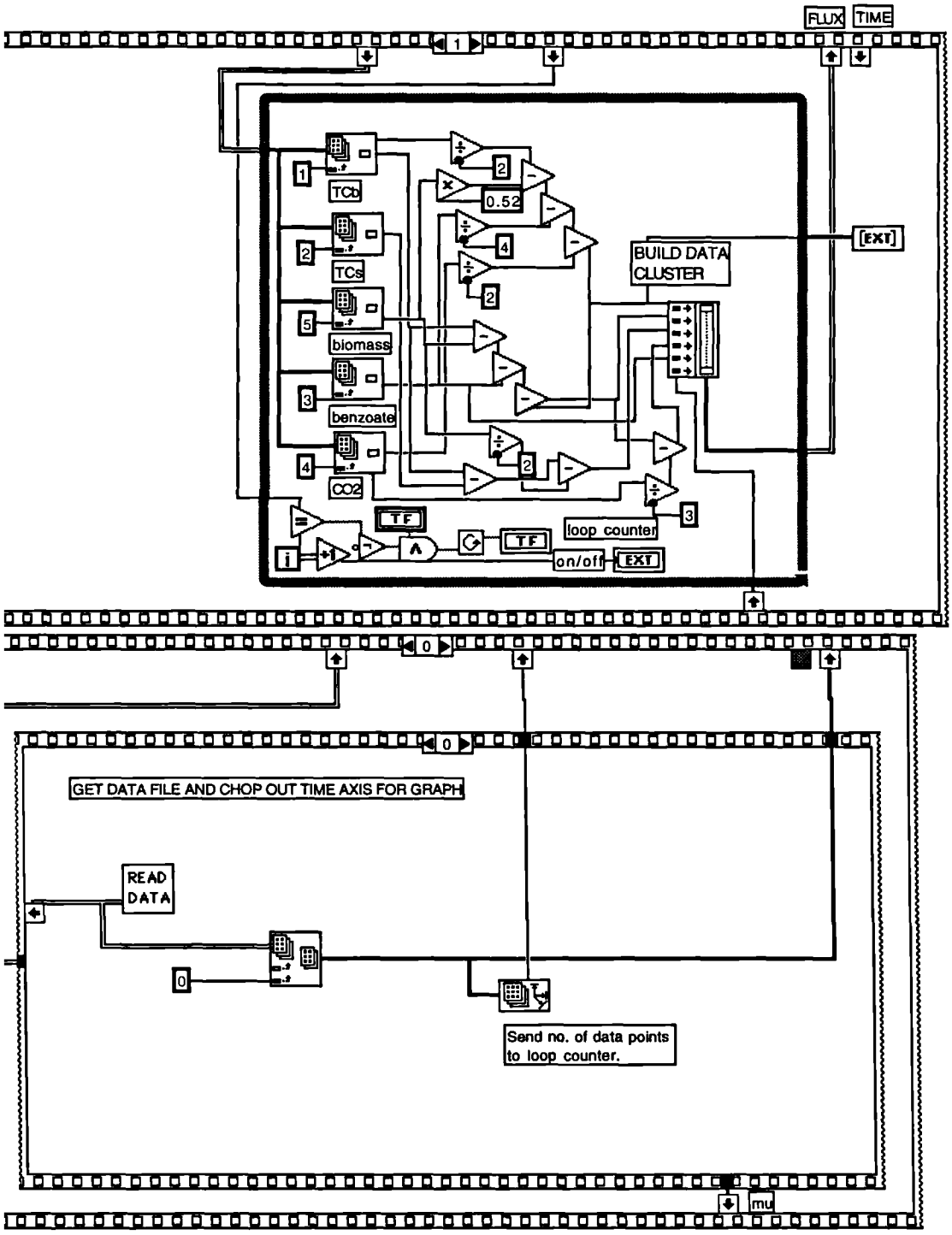
FLUX
CALC.

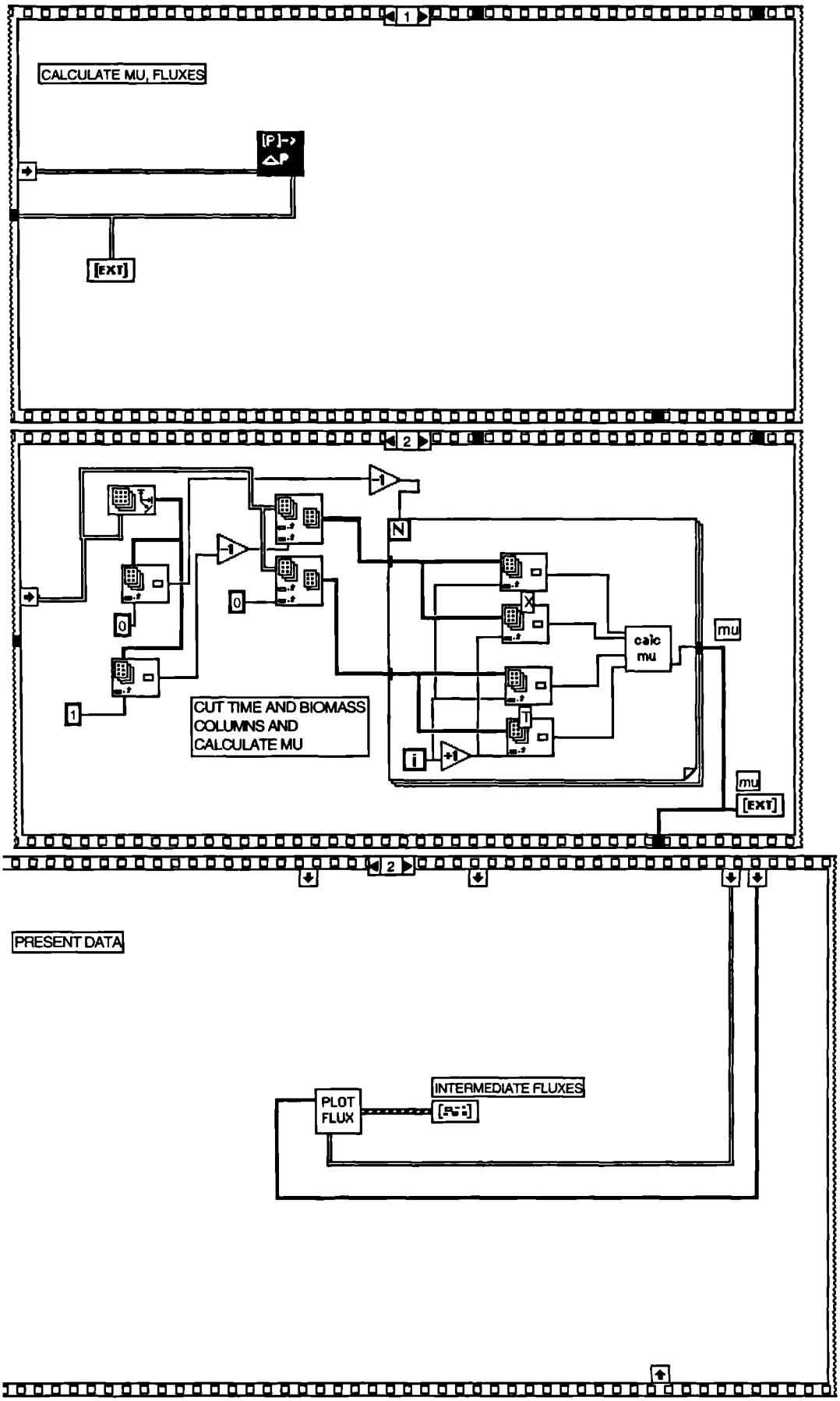
arbon Fluxes

Panel



jxes
23:19
jram




1
CALCULATE MU, FLUXES
[P]->
[EXT]
2
CUT TIME AND BIOMASS COLUMNS AND CALCULATE MU
N
calc mu
mu
[EXT]
2
PRESENT DATA
PLOT FLUX
INTERMEDIATE FLUXES

BIBLIOGRAPHY

S.J. Assinder, P.A. Williams (1990) The Tol Plasmids - Determinants Of The Catabolism Of Toluene And The Xylenes. *Adv.Mic.Phys.*,**31**,1-69.

M.-A. Abril, C. Michan, K.N. Timmis, J.L. Ramos (1989) Regulator and Enzyme Specificities of the TOL Plasmid-Enclosed Upper Pathway for Degradation of Aromatic Hydrocarbons And Expansion of the Substrate Range of the Pathway., *J.Bacteriol.*,**171**, 6782-6790.

S. Aiba, M. Matsuoka (1979) Identification of Metabolic Model: Citrate Production from Glucose by *Candida lipolytica*. *Biotechnol.Bioeng.*,**21**,1373-1386.

K.R. Albe, M.H. Butler, B.E. Wright (1990) Cellular Concentrations of Enzymes and Their Substrates. *J.theor.Biol.*,**143**,163-195.

W. Babel (1990) The Mixed Substrate Concept Applied for Microbial Synthesis of Metabolites. *Biotech.Adv.*,**8**,261-275.

K. Backman, M.J. O'Connor, A. Maruya, E. Rudd, M. Radjai, V. DiPasquantonio, D. Shoda, R. Hatch *et al* (1990) Genetic Engineering of Metabolic Pathways Applied to Phenylalanine. *Ann.N.Y.Acad.Sci.Biochem.Eng.*VI,16-90.

J.E. Bailey (1991) Toward a Science of Metabolic Engineering. *Science*,**252**,1668-1675.

J.E. Bailey, D.D Axe, P.M. Doran, J.L. Galazzo, K.F. Reardon, A. Seressiotis, J.V. Shanks (1987) Redirecton of Cellular Metabolism:Analysis and Synthesis. *Ann.N.Y. Acad.Sci.Biochem.Eng.*V,506,1-23.

J.E. Bailey, S. Birnbaum, J.L. Galazzo, C. Khosla, J.V. Shanks (1990) Strategies and Challenges in Metabolic Engineering. *Annals N.Y.Acad.Sci.Biochem.Eng.*VI,1-15.

J.E. Bailey, D.F. Ollis (1987) Biochemical Engineering Fundamentals. McGraw-Hill Intl. Editions

J.E. Bailey, J.V. Shanks (1991) ^{31}P NMR and ^{13}C NMR Studies of Recombimant *Saccheromyces cerevisiae* with Altered Glucose Phosphorylation Activities. *Bioprocess. Eng.*,**6**,273-284.

R.C. Benjamin, J.A. Voss, D.A. Kunz (1991) Nucleotide-Sequence Of XylE From The Tol Pdk1 Plasmid And Structural Comparison With Isofunctional Catechol-2,3- Dioxygenase Genes From Tol PWW0 And NAH7. *J.Bact.*,**173**,2724-2728.

P. Bery (1991) Reverse Genetics: Its Origins and Prospects. *Bio/Technology*,**9**,342-344.

T.H. Bley (1990) State-Structure Models - A Base for Efficient Control of Fermentation Processes. *Biotech.Adv.*,**8**,233-259.

J. Bonnerjea, J. Jackson, M. Hoare, P. Dunhill (1988) Affinity Flocculation of Yeast Cell Debris by Carbohydrate Specific Compounds. *Enzyme Microb.Technol.*, **10**, 357 -361.

J. Bonnerjea, S. Oh, M. Hoare, P. Dunhill (1986) Protein Purification: the Right Step at the Right Time. *Biotechnology*,**4**,954-956.

D. Bonvin, D.W.T. Rippin (1990) Target Factor Analysis for the Identification of Stoichiometric Models. *Chem.Eng.Sci.*,**45**,3417-3426.

E.J. Bormann, H.H. Grosse, J. Menz (1990) Entropy Balances of Microbial Product Formation. *Biotech.Adv.*,**8**,277-290.

J. Boudrant (1990) Microbial Processes for Ascorbic Acid Biosynthesis: A Review. *Enzyme Microb.Technol.*,**12**,322-329.

U. Brinkmann, W. Reineke (1992) Degradation Of Chlorotoluenes By In Vivo Constructed Hybrid Strains - Problems Of Enzyme Specificity, Induction And Prevention Of Meta-Pathway. *FEMS Microbiology Letters*,**96**,81-87.

W.J. Van der Broek, J. Visser (1986) Structure and Regulation of the Pyruvate Kinase Gene of *Aspergillis nidulans* L. de Graff. *Heredity*,**57**,130-130.

G.C. Brown, R.P. Hafner, M.D. Brand (1990) A "Top-Down" Approach to the Determination of Control Coefficients in Metabolic Control Theory. *Eur.J.Biochem.*, **188**,321-325.

T.D.K. Brown, M.C. Jones-Mortimer, H.L. Kornberg (1977) The Enzymic Interconversion of Acetate and Acetyl-coenzyme A in *Escherichia coli*. *J.Gen. Microbiol.*,**102**,327-336.

R.S. Burlage, S.W. Hooper, G.S. Sayler (1989) Minireview: The TOL(pWWO) Catabolic Plasmid. *Appl.Environ.Microbiol.*,**55**,132-136.

M.E. Bushell (1990) The Process Physiology of Secondary Metabolite Production.

S. Calero, E. Gari, I. Gilbert, J. Barbe (1989) Expression of the meta-Cleavage Pathway Operon of the TOL Plasmid of *Pseudomonas putida* in the Phototrophic Bacterium *Rhodobacter sphaeroides*. *J.Biotechnol.*,**12**,231-246.

E.I. Canela, R. Franco, M. Cascante (1989) Interdependence Between Cooperativity and Control Coefficients. *Biosystems*,**23**,7-14.

J.W. Cary, D.J. Petersen, G.N. Bennett, E.T. Papoutsakis (1990) Methods for Cloning Key Primary Metabolic Enzymes and Ancillary Proteins Associated with the Acetone-Butanol Fermentations of *Clostridium acetobutylicum*. *Annals N.Y.Acad. Sci.Biochem.Eng.*VI.67-81.

M. Cascante, R. Franco, E.I. Canela (1989a) Use of Implicit Methods from General Sensitivity theory to Develop a Systematic Approach to Metabolic Control. I. Unbranched Pathways. *Math.Biosci.*,**94**,271-288.

M. Cascante, R. Franco, E.I. Canela (1989b) Use of Implicit Methods from General Sensitivity Theory to Develop a Systematic Approach to Metabolic Control. II. Complex Systems. *Math.Biosci.*,**94**,289-309.

R.M Chalmers, J.N. Keen, C.A. Fewson (1991) Comparison Of Benzyl Alcohol Dehydrogenases And Benzaldehyde Dehydrogenases From The Benzyl Alcohol And Mandelate Pathways In Acinetobacter-Calcoaceticus And From The Tol-Plasmid-Encoded Toluene Pathway In Pseudomonas-Putida - N-Terminal Amino-Acid-Sequences, Amino-Acid Compositions And Immunological Cross-Reactions. *Biochem.J.*,**273**,99-107.

L.T. Chang, E.L. McGrory, R.P. Elander (1990) Penicillin Production by Glucose-Derepressed Mutants of *Penicillum chrysogenum*. *J.Ind.Microbiol.*,**6**,165-169.

W.-B.Z. Chu, A. Constantinides (1988) Modeling,Optimization and Computer Control of the Cephalosporin C Fermentation Process. *Biotechnol.Bioeng.*,**32**,277-288.

Y.B. Choi, J.Y. Lee, H.S. Kim (1992) A Novel Bioreactor For The Biodegradation Of Inhibitory Aromatic Solvents - Experimental Results And Mathematical-Analysis. *Biotechnol.Bioeng.*,**40**,1403-1411.

P.A.M. Claassen, C. Dijkema, J. Visser, A.J.B. Zehnder (1986) In vivo ^{13}C NMR Analysis of Acetate Metabolism in *Thiobacillus versutus* Under Denitrifying Conditions. *Arch.Microbiol.*,**146**,227-232.

D.P. Clark (1989) The Fermentation Pathways of *Escherichia coli*. *FEMS Microbiol. Rev.*,**63**,223-234.

F. Claveriemartin, B. Magasanik (1992) Positive And Negative Effects Of Dna Bending On Activation Of Transcription From A Distant Site. *J.Mol.Biol.*,**227**,996-1008.

W.L. Collinsworth, P.J. Chapman, S. Dagley (1973) Stereospecific Enzymes in the Degradation of Aromatic Compounds of *Pseudomonas putida*. *J.Bact.*,**113**,922-931.

J. Cordewener, R. Busink, J. Visser (1989) A Permeabilized Cell Assay System for Studying Enzyme Regulation and Localization in *Aspergillis niger*. *J.Microb.Methods* ,**10**,231-240.

A. Cornish-Bowden (1989) Metabolic Control Theory and Biochemical Systems Theory: Different Objectives, Different Assumptions, Different Results *J.theor.Biol.*,**136**, 365-377.

B. Crabtree, E.A. Newsholme (1988) Control Analysis Based on Unvalidated Models Cannot Be Used as a Short-Cut to Understanding Metabolic Control. *Biochem.J.*,**253**, 620-621.

B. Crabtree, E.A. Newsholme (1987a) A Systematic Approach to Describing and Analysing Metabolic Control Systems. *TIBS*,**12**,4-12.

B. Crabtree, E.A. Newsholme (1987b) The Derivation and Interpretation of Control Coefficients. *Biochem.J.*,**247**,113-120.

D.L. Cruden, J.H. Wolfram, R.D. Rogers, D.T. Gibson (1992) Physiological- Properties Of A Pseudomonas Strain Which Grows With Paraxylene In A 2-Phase (Organic-Aqueous) Medium. *Appl.Env.Microbiol.*,**58**,2723-2729.

S. Dagley (1986) Biochemistry of Aromatic Hydrocarbon Degradation in Pseudomonas. The Bacteria:Vol X The Biology of the Pseudomonas pp 527-555.

K. Van Dam, M.M. Mulder, J. Teixeira de Mattos, H.V. Westerhoff (1988) A Thermodynamic View of Bacterial Growth. Physiological Models in Microbiology pp 25-48. Ed.s M.J. Bazin, J.I. Prosser CRC Press.

A. Delgado, M.G. Wubbolts, M.A. Abril, J.L. Ramos (1992) Nitroaromatics Are Substrates For The Tol Plasmid Upper-Pathway Enzymes. *Appl.Env.Microbiol.*,**58**, 415-417.

J. Delgado, J.C. Liao (1992) Determination of Flux Control Coefficients from Transient Metabolite Concentrations. *Biochem.J.*,**282**,919-927.

J.P. Delgado, J.C. Liao (1991) Identifying Rate-Controlling Enzymes in Metabolic Pathways Without Kinetic Parameters. *Biotechnol.Prog.*,**7**,15-20.

V. Delorenzo, M. Herrero, M. Metzke, K.N. Timmis (1991) An Upstream Xylr-Induced And Ihf-Induced Nucleoprotein Complex Regulates The Sigma-54-Dependent Pu Promoter Of Tol Plasmid. *EMBO J.*,**10**,1159-1167.

R.F. Derr (1986) Modern Metabolic Control Theory: II Determination of Flux Control Coefficients.

Biochem.Arch.,2,31-44.

G.A. Dervakos, J.M. Woodley, E.T. Keravnou, J. Washbrook, M.D. Lilly (1989) Development of a KBS for Biotransformation Process Design. *I.Chem.E.Symp.Series* No.**114**,283-292.

J.C. Diaz-Ricci, B. Hitzmann, U. Rinas, J.E. Bailey (1990) Comparative Studies of Glucose Cataboism by *E.coli* Grown in a Complex Medium Under Aerobic and Anaerobic Conditions. *Biotechnol.Prog.*,**6**,326-332.

J.-C. Diaz-Ricci, M. Tsu, J.E. Bailey (1992) Influence of Expression of the pet Operon on Intracellular Metabolic Fluxes of *Escherichia coli*. *Biotechnol.Bioeng.*,**39**,59-65.

C. Dijkema, R.P. Rijcken, H.C.M. Kester, J. Visser (1986) ^{13}C NMR Studies on the Influence of pH and Nitrogen Source on Polyol Pool Formation in *Aspergillus nidulans*. *FEMS Microbiol.Letters*,**33**,125-131.

C. Dijkema, J. Visser (1987) ^{12}C NMR Analysis of Aspergillus Mutants Disturbed in Pyruvate Metabolism. *Biochim.Biophys.Acta.*,**931**,311-319.

G. Dodson (1989) Realistic Protein Engineering;Site-Directed Mutagenesis;A Review. *Soc.Gen.Microbiol.Symp.*,**44**,293-307.

W.A. Duetz, J.G. Vanandel (1991) Stability Of Tol Plasmid Pww0 In *Pseudomonas Putida* Mt-2 Under Nonselective Conditions In Continuous Culture. *J.Gen.Microbiol.*, **137**,1369-1374.

W.A. Duetz, M.K. Winson, J.G. Vanandel, P.A. Williams (1991) Mathematical Analysis Of Catabolic Function Loss In A Population Of *Pseudomonas Putida* Mt-2 During Non-Limited Growth On Benzoate. *J.Gen.Microbiol.*,**137**,1363-1368.

C.J. Duggleby, P.A. Williams (1986) Purification and Some Properties of the 2-Hydroxy-6-oxahepta-2,4-dienoate Hydrolase (2-Hydroxymuconic Semialdehyde Hydrolase) Encoded by the TOL Plasmid pWWO from *Pseudomonas putida* mt-2. *J.Gen.Microbiol.*,**132**,717-726.

A. Earl (1991) The Genetic Engineering Approach to B-Lactam Antibiotic Stain Improvement. *J.Chem.Tech.Biotechnol.*,**50**,123-126.

J.S. Easterby (1990) Integration of Temporal Analysis and Control Analysis of Metabolic Systems. *Biochem.J.*,**269**,255-259.

B.D. Ensley, D.T. Gibson (1983) Naphthalene Dioxygenase: Purification and Properties of a Terminal Oxygenase Component. *J.Bacteriol.*,**155**,505-511.

D.A. Fell (1992) Metabolic Control Analysis: a Survey of its Theoretical and Experimental Development. *Biochem.J.*,**286**,313-330.

D.A. Fell, H.M. Sauro (1985) Metabolic Control and its Analysis. *Eur.J.Biochem.*, **148**,555-561.

D.A. Fell, H.M. Sauro (1990) Metabolic Control Analysis: The Effects of high Enzyme Concentrations. *Eur.J.Biochem.*,**192**,183-187.

S. Fetzner, R. Muller, F. Lingens (1992) Purification And Some Properties Of 2-Halobenzoate 1,2-Dioxygenase, A 2-Component Enzyme-System From *Pseudomonas Cepacia* 2cbs. *J.Bact.*,**174**,279-290.

J.M. Foght, D.W.S. Westlake (1991) Cross Hybridization Of Plasmid And Genomic DNA From Aromatic

And Polycyclic Aromatic Hydrocarbon Degrading Bacteria. *CanJ.Microbiol.*,**37**,924-932.

F.C.H. Franklin, M. Bagdasarian, M.M. Bagdasarian, K.N. Timmis (1981) Molecular and Functional Analysis of the TOL Plasmid pWWO from *Pseudomonas putida* and Cloning of Genes for the Entire Regulated Aromatic Ring meta Cleavage Pathway. *Proc.Natl.Acad.Sci.*USA,**78**,111-121.

B. Frantz, A.M. Chakrabarty (1986) Degradative Plasmids in Pseudomonas. The Bacteria:Vol X The Biology of the Pseudomonas pp 295-323. Ed.s J.R. Sokatch, L.N. Ornston

J.K. Fredrickson, F.J. Brockman, D.J. Workman, S.W. Li, T.O. Stevens (1991) Isolation And Characterization Of A Subsurface Bacterium Capable Of Growth On Toluene, Naphthalene, And Other Aromatic Compounds. *Appl.Env.Microbiol.*,**57**, 796-803.

L. Fryklund, P. Gellerfors, K. Fholenhag, B. Holstrom (1989) Recombinant Proteins as Pharmaceutical Products: The Interplay Between Downstream Procesing and Protein Characterization. *Soc.Gen.Microbiol.Symp.*,**44**,187-193.

K. Furukawa, N. Hayase, K. Taira, N. Tomizuka (1989) Molecular Relationship of Chromosomal Genes Encoding Biphenyl/Polychlorinated Biphenyl Catabolism: Some Soil Bacteria Possess a Highly Conserved bph Operon. *J.Bacteriol.*,**171**,5467-5472.

K. Furukawa, S. Hayashida, K. Taira (1991) Gene-Specific Transposon Mutagenesis of the Biphenyl Polychlorination Biphenyl-Degradation-Controling bph operon in Soil Bacteria. *Gene*,**98**,21-28.

K. Furukawa, J. Hirose, A. Suyama, T. Zaiki, D.T. Gibson, S. Hayashida (1993) Gene Components Responsible for Discrete Subatrate Specificity in the Catabolosm of Biphenyl (bph Operon) and Toluene (tod Operon). *J.Bacteriol.*-submitted

K. Furukawa, H. Suzuki (1988) Gene Manipulation of Catabolic Activities for Production of Intermediates of Various Biphenyl Compounds. *Appl.Microbiol. Biotechnol.*,**29**,363-369.

J.L. Galazzo, J.V. Shanks, J.E. Bailey (1990) Comparison of Intracellular Sugar-Phosphate Levels from ^{31}P NMR Spectroscopy of Intact Cells and Cell-Free Extracts. *Biotech.Bioeng.*,**35**,1164-1168.

N. Gautan (1988) Mutated Forms of Phosphoglycerate Mutase in Yeast Affect Reversal of Metabolic Flux. *J.Biol.Chem.*,**263**,15400-15406.

D.T. Gibson, D.L. Cruden, J.D. Haddock, G.J. Zylstra, J.M. Brand (1993) Oxidation of Polychlorinated Biphenyls by *Pseudomonas* sp. Strain LB400 and *Pseudomonas pseudoalcaligenes* Strain KF707. submitted.

C. Giersch (1988a) Control Analysis of Metabolic Networks. 1. Homogenous Functions and the Summation Theorems for Control Coefficients. *Eur.J.Biochem.*,**174**,509-513.C. Giersch (1988b) Control Analysis of Metabolic Networks. 2. Total Differentials and General Formulation of the Connectivity Relations. *Eur.J.Biochem.*,**174**,515-519.

C. Giersch (1988c) Control Analysis of Biochemical Pathways: A Novel Procedure for Calculating Control Coefficients and an Additional Theorem for Branched Pathways. *J.theor.Biol.*,**134**,451-462.

I.E. Gleiser, S. Bauer (1981) Growth of *E.coli* W to High Cell Concentration by Oxygen Level Linked

Control of Carbon Source Concentration. *Biotechnol.Bioeng.*, **23**,1015-1021.

A. Goldbeter, G. Nicolis (1976) An Allosteric Enzyme Model with Positive Feedback Applied to Glycolytic Oscillations. *Prog.theor.Biol.*,**4**,65-160.

M. Gomada, S. Inouye, H. Imaishi, A. Nakazawa, T. Nakazawa (1992) Analysis Of An Upstream Regulatory Sequence Required For Activation Of The Regulatory Gene XylS In Xylene Metabolism Directed By The Tol Plasmid Of *Pseudomonas Putida. Mol.Gen.Genet.*,**233**,419-426.

A.K. Groen, R. Van Der Meer, H.V. Westerhoff, R.J.A. Wanders, T.P.M. Akerboom, J.M. Tager (1982) Control of Metabolic Fluxes. Metabolic Compartmentation pp 9-37. Ed. H. Sies.

A.K. Groen, J.M. Tager (1988) Control Analysis Provides a Simple Means of Understanding the Control Structure of a Metabolic Pathway. *Biochem.J.*,**253**,619-620.

E.E. Groseclose (1978) The Catabolism of Resorcinol by *Azotobacter vinelandii.* PhD Thesis, U. of Miami

K. Han, H.C. Lim, J. Hong (1992) Acetic Acid Formation in *Escherichia coli* Fermentation. *Biotechnol.Bioeng.*,**39**,663-671.

S. Harayama, R.A. Leppik, M. Rekik, N. Mermod, P.R. Lehrbach, W. Reineke, K.N. Timmis (1986) Gene Order of the TOL Catabolic Plasmid Upper-Pathway Operon and Oxidation of Both Toluene and Benzyl Alcohol by the xylA Product. *J.Bact.*,**167**, 455-461.

S. Harayama, N. Mermod, M. Rekik, P.R. Lehrbach, K.N. Timmis (1987) Roles of the Divergent Branches of the meta-Cleavage Pathway in the Degradation of Benzoate and Substituted Benzoates. *J.Bact.*,**169**,558-564.

S. Harayama, A. Polissi, M. Rekik (1991) Divergent Evolution Of Chloroplast-Type Ferredoxins. *FEBS Letters*,**285**,85-88.

S. Harayama, M. Rekik (1990) The meta Cleavage Operon of TOL Degenerative Plasmid pWWO Comprises 13 Genes. *Mol.Gen.Genet.*,**221**,113-120.

S. Harayama, M. Rekik (1989) A Simple Proceedure for Transferring Genes Cloned in *Escherichia coli* Vectors into Other Gram-Negative Bacteria: Phenotypic Analysis And Mapping of TOL Plasmid xylK. *Gene*,**78**,19-27.

S. Harayama, M. Rekik, K.-L. Ngai, L.N. Ornston (1989) Physically Associated Enzymes Produce and Metabolize 2-Hydroxy-2,4-Dienoate, a Chemically Unstable Intermediate Formed in Catechol Metabolism via meta Cleavage in *Pseudomonas putida*

J.Bact.,**171**,6251-6258.

S. Harayama, M. Rekik, K.N. Timmis (1986) Genetic Analysis of a Relaxed Subatrate Specificity Aromatic Ring Dioxygenase, Toluate 1,2-Dioxygenase, Encoded by TOL Plasmid pWWO of *Pseudomonas putida. Mol.Gen.Genet.*,**202**,226-234.

S. Harayama, M. Rekik, M. Wubbolts, K. Rose, R.A. Leppik, K.N. Timmis (1989) Characterization of 5 Genes in the Upper-Pathway Operon of TOL Plasmid pWWO from Pseudomonas putida and Identification of the Gene Products. *J.Bact.*,**171**, 5048-5055.

S. Harayama, K.N. Timmis (1989) Catabolism of Aromatic Hydrocarbons by Pseudomonas. Genetics of Bacterial Diversity pp 151-174 Ed.s D.A. Hopwood, K.F.Charter Academic Press N.Y.

N. Hayase, K. Taira, K. Furukawa (1990) *Pseudommonas putida* KF715 bphABCD Operon Encoding Biphenyl and Polychlorinated Biphenyl Degradation: Cloning, Analysis, and Expression in Soil Bacteria. *J.Bacteriol.*,**172**,1160-1164.

R. Heinrich, S.M. Rapoport, T.A. Rapoport (1977) Metabolic Regulation and Mathematical Models. *Prog.Biophys.Molec.Biol.*,**32**,1-35.

R. Heinrich, T.A. Rapoport (1974) A Linear Steady-State Treatment of Enzymatic Chains: Critique of the Crossover Theorem and a General Procedure to Identify Interaction Sites with an Effector. *Eur J.Biochem.*,**42**,97-105.

R. Heinrich, T.A. Rapoport (1975) Mathematical Analysis of Multienzyme Systems: II Steady State and Transient Control. *Biosystems*,**7**,130-136.

F.K. Higson, D.D. Focht (1992) Degradation Of 2-Methylbenzoic Acid By *Pseudomonas Cepacia* MB2. *Appl.Env.Microbiol.*,**58**,194-200.

M. Hoare, P. Dunhill (1986b) Protein Processing-New Prospects. *The Chem.Eng.*, 39-41.

M. Hoare, P. Dunhill (1986a) Processing with Proteins-The Immediate Challenge. *The Chem.Eng.*, ,23-25.

A. Holtel, K.N. Timmis, J.L. Ramos (1992) Upstream Binding Sequences Of The XylR Activator Protein And Integration Host Factor In The XylS Gene Promoter Region Of The Pseudomonas TOL Plasmid. *Nuc.Acid.Res.*,**20**,1755-1762.

F. Hofer, E. Weissinger, H. Mischak, R. Messner, B. Meixner-Monori, D. Blaas, J. Visser, C.P. Kubicek (1989) A Monoclonal Antibody Against the Alkaline Extracellular b-Glucosidase from Trichoderma reesei: Reactivity with Other Trichoderma B-Glucosidases. *Biochim.Biophys.Acta.*,**992**,298-306.

J.H.S. Hofmeyr (1989) Control-Pattern Analysis of Metabolic Pathways: Flux and Concentration Control in Linear Pathways. *Eur.J.Biochem.*,**186**,343-354.

J.H.S. Hofmeyr (1986) Steady-State Modeling of Metabolic Pathways: A Guide for the Prospective Simulator. *CABIOS*,**2**,5-11.

W.H. Holmes (1986) The Central Metabolic Pathways of Escherichia coli: Relationship between Flux and Control ata Branch Point, Efficiency of Conversion to Biomass and Excretion of Acetate. *Curr.Top.Cell.Reg.*,**28**,69-105.

W.H. Holmes, I.D. Hamilton, D. Mousdale (1991) Improvements to Microbial Productivity by Analysis of Metabolic Fluxes. *J.Chem.Tech.Biotechnol.*,**50**,139-141.

Z. Hostalek, J. Novotna, V. Stary, L. Kalachova, J. Vorisek (1990) Subcellular Organisation of Streptomyces aurofaciens and Overproduction of Chlortetracycline. *Biotech.Adv.*,**8**,131-139.

J.M. Horn, S. Harayama, K.N. Timmis (1991) Dna-Sequence Determination Of The Tol Plasmid (Pwwo) XylGFJ Genes Of Pseudomonas-Putida - Implications For The Evolution Of Aromatic Catabolism. *Mol.Microbiol.*,**5**,2459-2474.

J. Huang, P. Dhulster, D. Thomas, J.-N. Barbotin (1990) Agitation Rate Effects on Plasmid Stability in

Immobilized and Free-Cell Continuous Cultures of Recombinant *E.coli*. *Enzyme Microb.Technol.*,**12**,933-939.

I.S. Hunter, S. Baumberg (1989) Molecular Genetics of Antibiotic Formation-Gene Cloning and Expression. *Soc.Gen.Microbiol.Symp.*,**44**,121-162.

C.R. Hutchinson (1992) Recombinant DNA and the Development of Antitumor and Other Antibiotecs Produced by Actinomycetes. *Pharmaceutical Tech.*, ,22-27.

A. Inoue, K. Honkoshi (1989) A Pseudomonas Thrives in High Concentrations of Toluene. *Nature*,**338**,264-266.

S. Inouye, A. Nakazawa, T. Nakazawa (1988) Nucleotide Sequence of the Regulatory Gene xylR of the TOL Plasmid from Pseudomonas putida. *Gene*,**66**,301-306.

S. Inouye, A. Nakazawa, T. Nakazawa (1986) Nucleotide Sequence of the Regulatory Gene xylS on the Pseudomonas putida TOL Plasmid and Identification of the Protein Product. *Gene*,**44**,235-242.

S. Inouye, A. Nakazawa, T. Nakazawa (1984) Nucleotide Sequence of the Promoter Region of the xylDEGF Operon on the TOL Plasmid of Pseudomonas putida. *Gene*,**29**, 323-330.

S. Irie, S. Doi, T. Yorifuji, T. Masamichi, K. Yano (1987) Nucleotide Sequencing and Characterization of the Genes Encoding Benzene Oxidation Enzymes of Pseudomonas putida. *J.Bact.*,**169**,5174-5179.

D.H. Irvine (1990) Objectives, Assumptions and Results of Metabolic Control Theory and Biochemical Systems Theory. *J.theor.Biol.*,**143**,139-143.

R.K. Jain, R.C. Bayly, R.A. Skurray (1991) Specific Deletion Of A Large Segment Of Pra500 - A 3,5-Xylenol Degradative Plasmid. *Lett.Appl.Microbiol.*,**12**,216-220.

W.H. Jeffrey, S.M. Cuskey, P.J. Chapman, S. Resnick, R.H. Olsen (1992) Characterization Of Pseudomonas-Putida Mutants Unable To Catabolize Benzoate - Cloning And Characterization Of Pseudomonas Genes Involved In Benzoate Catabolism And Isolation Of A Chromosomal Dna Fragment Able To Substitute For XylS In Activation Of The TOL Lower-Pathway Promoter. *J.Bacteriol.*,**174**,4986-4996.

R.O. Jenkins, G.M. Stephens, H. Dalton (1987) Production of Toluene cis-Glycol by *Pseudomonas putida* in Glucose Fed-Batch Culture. *Biotechnol.Bioeng.*,**29**,873-883.

J. Jeong, M.M. Ataai (1990) A Highly Structured Model for Simulation of Batch and Continuous Cultures of *Bacillus subtilis* and Examination of Cellular Differentation. *Ann.N.Y.Acad.Sci.Biochem.Eng.***VI**,82-90.

C.W. Jones (1991) Energy Transduction and Microbial Performance. *J.Chem.Tech. Biotechnol.*,**50**,131-133.

G. Jung, P. Denefle, J. Becquart, J.-F. Mayaux (1988) High-Cell Density Fermentation Studies of Recombinant *Escherichia coli* Strains Expressing Human Interleukin-1b. *Ann.Inst.Pasteur/Microbiol.*,**139**,129-146.

H. Kacser, J.A. Burns (1973) The Control of Flux. *Symp.Soc.exp.Biol.*,**27**,65-104.

H. Kacser, J.W. Porteus (1987) Control of Metabolism: What Do We Have To Measure? *TIBS*,**12**,5-13.

D. Kahn, H.V. Westerhoff (1991) Control Theory of Regulatory Cascades. *J.theor. Biol.*,**153**,255-285.

P.D. Keightly (1989) Models of Quantative Variation of Flux in Metabolic Pathways.

Genetics,**121**,869-876.

S. Keil, H. Keil (1992) Construction Of A Cassette Enabling Regulated Gene Expression In The Presence Of Aromatic Hydrocarbons. *Plasmid*,**27**,191-199.

T. Keleti, J. Ovadi (1988) Control of Metabolism by Dynamic Macromolecular Interactions. *Curr.Top.Cell.Reg.*,**29**,1-33.

D.B. Kell, K Van Dam, H.V. Westerhoff (1989) Control Analysis of Microbial Growth Productivity. *Soc.Gen.Microbiol.Symp.*,**44**,61-93.

D.B. Kell, H.V. Westerhoff (1986a) Metabolic Control Theory: Its Role in Microbiology and Biotechnology. *FEMS Microbiol.Rev.*,**39**,305-320.

D.B. Kell, H.V. Westerhoff (1986b) Towards a Rational Approach to the Optimization of Flux in Microbial Biotransformations. *TIBTECH*, ,137-142.

D.B. Kell, H.V. Westerhoff (1990) Control Analysis of Organized Multienzyme Systems. Structural and Organizational Aspects of Metabolic Regulation. B.N. Kholodenko (1988) How Do External Parameters Control Fluxes and Concentrations of Metabolites? An Additional Theorem in the Theory of Metabolic Control. *FEBS Letters*,**232**,383-386.

A. Kiener (1992) Enzymatic Oxidation Of Methyl-Groups On Aromatic Heterocycles - A Versatile Method For The Preparation Of Heteroaromatic Carboxylic-Acids. *Angewandte Chemie-International (English)*,**31**,774-775.

Y. Kim, B. Choi, J. Lee, H. Chang, K.R. Min (1992) Characterization Of Catechol 2,3-Dioxygenases. *Biochem.Biophys.Res.Comm.*,**183**,77-82.

R.D. Kiss, G. Stephanopoulos (1992) Metabolic Characterization of a L-Lysine Producing Strain by Continuous Culture. *Biotechnol.Bioeng.*,**39**,565-574.

H. Kleinkauf, H. von Dohren (1990) Antibiotics - Cloning of Biosynthetic Pathways. *FEBS Letters*,**268**,405-409.

M.C. Kohn, E. Chiang (1982) Metabolic Network Sensitivity Analysis. *J.theor.Biol.*, **98**,109-126.

M.C. Kohn, D.R. Lemieux (1991) Identification of Regulatory Properties of Metabolic Networks by Graph Theoretical Modelling. *J.theor.Biol.*,**150**,3-25.

M.C. Kohn, L.M. Whitley, D. Garfinkel (1979) Instantaneous Flux Control Analysis for Biological Systems. *J.theor.Biol.*,**76**,437-452.

K.B. Konstantinov, T. Yoshida (1992) Mini-Review: Knowledge-Based Control of Fermentation Processes. *Biotechnol.Bioeng.*,**39**,479-486.

D.A. Kunz, P.A. Chapman (1981) Catabolism of Pseudocumene and 3-Ethyltoluene by Pseudomonas putida (arvilla) mt-2: Evidence for New Functions of the TOL Plasmid. *J.Bacteriol.*,**146**,179-191.

K.A. Lampel (1979) The p-Cymene pathway in Pseudomonas putida: a Comparative Study of the Three Enzymes Catalyzing Sequential Reactions Subsequent to Ring Cleavage. PhD Thesis, U. of Miami

R.A. Lazarus, J.L. Seymour, R.K. Stafford, M.S. Dennis, M.G. Lazarus, C.B. Marks, S. Anderson (1990) A Biocatalytic Approach to Vitamin C Production - Metabolic Pathway Engineering of Erwina herbicola.

Biocatalysis Ed. D.A. Abramowicz P.R. Lehrbach, I. McGregor, J.M. Ward, P. Broda (1983) Molecular Relationships Between Pseudomonas INC P-9 Degradative Plasmids TOL, NAH and SAL. *Plasmid*, **10**,164-174.

P.R. Lehrbach, J. Zeyer, W. Reineke, H.-J. Knackmuss, K.N. Timmis (1984) Enzyme Recruitment In Vitro: Use of Cloned Genes to Extend the Range of Haloaromatics Degraded by Pseudomonas sp. Strain B13. *J.Bacteriol.*,**158**,1025-1032.

S.M. LeVine, F. Ardeshir, G.F.-L. Ames (1980) Isolation and Characterization of Acetate Kinase and Phosphotransacetylase Mutants of *Escherichia coli* and *Salmonella typhimurium*. *J.Bact.*,**143**,1081-1085.

M.D. Lilly, J.D. Woodley (1985) Biocatalytic Reactions Involving Water-Insoluble Organic Compounds. Biocatalysis in Organic Synthesis pp.179-192 Ed.s J.Tramper, H.C.van der Plas, Elsevier

G.W. Lubi, W.R. Strohl (1990) Comparison of Growth, Acetate Production and Acetate Inhibition of *E.coli* Strains in Batch and Fed-Batch Fermentations. *Appl.Env.Microbiol.*,**56**,1004-1011.

J. Lobley, H. Keil (1991) Homologies Between Plasmid And Chromosomally-Encoded Benzoate-Oxidizing Genes In *Pseudomonas Putida*. *Lett.Appl.Microbiol.*,**13**,66-70.

G. Lybertos, B. Kuszta, J.E. Bailey (1985) Versal Matrix Families, Normal Forms and Higher Order Bifurcations in Dynamic Chemical Systems. *Chem.Eng.Sci.*,**40**, 1177-1189.

G. Lybertos, B. Kuszta, J.E. Bailey (1984) Steady-State Multiplicity and Bifurcation Analysis via the Newton Polyhedra Approach. *Chem.Eng.Sci.*,**39**,947-960.

A.A. Mae, Y.K. Habicht, A.E. Nurk, A.L. Heinaru (1991) Transposons Tn4652 And Tn3614 Of The Tol Plasmid PWW0 Are Involved In Conjugal Mobilization Of Chromosomally Located Catabolic Cam-Operons. *Genetika*,**27**,773-782.

R.A. Majewski, M.M. Domach (1990) Effect of Regulatory Mechanism on Hyperbolic Reaction Network Properties. *Biotechnol.Bioeng.*,**36**,166-178.

L.H. Malmberg, W.-S. Hu (1992) Identification of Rate-Limiting Steps in Cephalosporin C Biosynthesis in Cephalosporium acremonium: a Theoretical Anaysis. *Appl.Microbiol.Biotechnol.*,**38**,122-128.

L.-H. Malmberg, W.-S. Hu (1991) Kinetic Analysis of Cephalosporin Biosynthesis in *Streptomyces clavuligarus*. *Biotech.Bioeng.*,**38**,941-947.

J.L. Marx (1989) The Cell Cycle Coming Under Control. *Science*,**245**,252-255.

A. Matsuyama, H. Yamamoto, E. Nakano (1989) Cloning, Expression, and Nucleotide Sequence of the *Escherichia coli* K-12 ackA Gene. *J.Bact.*,**171**,577-580.

M.L. Mavrovouniotis, G. Stephanopoulos, G. Stephanopoulos (1989) Formal Modeling of Approximate Relations in Biochemical Systems. *Biotechnol.Bioeng.*,**34**,196-206.

E. Melendez-Hevia, N.V. Torres (1988) Economy of Design in Metabolic Pathways. Further Remarks on the Game of the Pentose Phosphate Cycle. *J.theor.Biol.*, **132**,97-111.

E. Melendez-Hevia, N.V. Torres, J. Sicilia, H. Kacser (1990) Control Analysis of Transition Times in Metabolic Systems. *Biochem.J.*,**265**,195-202.

E. Mendelez-Hevia, N.V. Torres, J. Sicilia (1990) A Generalization of Metabolic Control Analysis to

Conditions of No Proportionality Between Activity and Concentration of Enzymes. *J.theor.Biol.*,**142**,443-451.

N. Mermod, S. Harayama, K.N. Timmis (1986) New Route to Bacterial Production of Indigo. *Biotechnology*,**4**,321-324.

N. Mermod, P.R. Lehrbach, R.H. Don, K.N. Timmis (1986) Gene Cloning and Manipulation in Pseudomonas. The Bacteria:Vol X The Biology of the Pseudomonas pp 325-355 Ed.s J.R. Sokatch, L.N. Ornston

H.-P. Meyer, C. Leist, A. Feichter (1984) Acetate Formation in Continuous Culture of *Escherichia coli* K12 D1 on Defined and Complex Media. *J.Biotechnol.*,**1**,355-358.

C. Michan, B. Kessler, V. Delorenzo, K.N. Timmis, J.L Ramos (1992) XylS Domain Interactions Can Be Deduced From Intraallelic Dominance In Double Mutants Of *Pseudomonas putida*. *Mol.Gen.Genet.*,**235**,406-412.

C. Michan, L.M. Zhou, M.T. Gallegos, K.N. Timmis, J.L. Ramos (1992) Identification Of Critical Amino-Terminal Regions Of XylS - The Positive Regulator Encoded By The TOL Plasmid. *J.Biol.Chem.*,**267**,22897-22901.

E. Mizraji, L. Acerenza, J. Hernandez (1988) Time Delays in Metabolic Control Systems. *Biosystems*,**22**,11-17.

Y. Murooka (1990) Genetic Designs for Product Formation in Recombinant Microbes. *Biotech.Adv.*,**8**,29-57.

K. Murray, C.J. Duggleby, J.M. Sala-Trepat, P.A. Williams (1972) The Metabolism of Benzoate and Methylbenzoates via the meta-Cleavage Pathway by *Pseudomonas arvilla* mt-2. *J.Biochem.*,**28**,301-310.

C. Nakai, K. Hori, H. Kagamiyama, T. Nakazawa, M. Nozaki (1983) Purification, Subunit Structure and Partial Amino Acid Strucyure of Metapyrocatechase. *J.Biol.Chem.*,**258**,2916-2922.

C. Nakai, H. Kagamiyama, M. Nozaki (1983) Complete Nucleotide Sequence of the Metapyrocatechase Gene on the TOL Plasmid of *Pseudomonas putida* mt-2. *J.Biol.Chem.*,**258**,2923-2928.

T. Nakazawa, S. Inouye (1986) Cloning of Pseudomonas Genes in *Escherichia coli*. The Bacteria:Vol X The Biology of the Pseudomonas pp 357-382 Ed.s J.R. Sokatch, L.N. Ornston.

C. Nakatsu, J. Ng, R. Singh, N. Straus, C. Wyndham (1991) Chlorobenzoate Catabolic Transposon Tn5271 Is A Composite Class-I Element With Flanking Class-II Insertion Sequences. *Proc.Nat.Acad.Sci.U.S.A.*,**88**,8312-8316.

F.C. Neidhardt, P.L. Bloch, D.F. Smith (1974) Culture Medium for Enterobacteria. *J.Bact.*,**119**,736-747.

E. Neidle, C. Hartnett, L.N. Ornston, A. Bairoch, M. Rekik, S. Harayama (1992) Cis-Diol Dehydrogenases Encoded By The Tol Pwwo Plasmid XylL-Gene And The Acinetobacter-Calcoaceticus Chromosomal Bend-Gene Are Members Of The Short-Chain Alcohol-Dehydrogenase Superfamily. *Eur.J.Biochem.*,**204**,113-120.

E.L. Neidle, M.K. Shapiro, L.N. Ornston (1987) Cloning and Expression in *Escherichia coli* of *Acinetobacter calcoaceticus* Genes for Benzoate Degradation. *J.Bacteriol.*,**169**,5496-5503.

J. Neilsen, K. Nikolajsen, J. Villadsen (1991) Structured Modelling of a Microbial System: I. A Theoretical Study of Lactic Acid Fermentation. *Biotechnol.Bioeng.*, 38,1-10.

J. Nielsen, C. Emborg, K. Halberg, J. Villadsen (1989) Compartment Model Concept Used in the Design of Fermentation with Recombinant Microorganisms. *Biotechnol. Bioeng.*,34,478-486.

J. Nielsen, K. Nikolajsen, J.Villadsen (1991b) Structured Modelling of a Microbial System: II. Experimental Verification of a Structured Lactic Acid Fermentation Model. *Biotechnol.Bioeng.*,38,11-23.

J. Nielsen, A.G. Pedersen, K. Strudsholm, J. Villadsen (1991) Modelling Fermentations with Recombinant Microorganisms: Formulation of a Structured Model. *Biotechnol.Bioeng.*,37,802-808.

H.G. Nimmo, A.C. Borthwick, E.M.T. El-Mansi, W.H. Holms, C. MacKintosh, G.A. Nimmo (1987) Regulation of the Enzymes at the Branchpoint between the Citric Acid Cycle and the Glyoxylate Bypass in *Escherichia coli.*

H.G. Nimmo, P.T.W. Cohen (1987) Applications of Recombinant DNA Technology to Studies of Metabolic Regulation. *Biochem.J.*,247,1-13.

S.C. Niranjan, K.Y. San (1989) Analysis of a Framework Using Material Balances in Metabolic Pathways to Elucidate Cellular Metabolism. *Biotech.Bioeng.*,34,496-501.

H.J. Noorman, J.J. Heijnen, K.Ch.A.M. Luyben (1991) Linear Systems in Microbial Reaction Systems: A General Overview of their Origin, Form, and Use. *Biotechnol. Bioeng.*,38,603-618.

K. Nusslein, D. Maris, K. Timmis, D.F. Dwyer (1992) Expression And Transfer Of Engineered Catabolic Pathways Harbored By Pseudomonas Spp Introduced Into Activated-Sludge Microcosms. *Appl.Env.Microbiol.*,58,3380-3386.

K.J. O'Donnell, P.A. Williams (1991) Duplication Of Both Xyl Catabolic Operons On Tol Plasmid Pww15. *J.Gen.Microbiol.*,137,2831-2838.

T. Omori, H. Ishigooka, Y. Minoda (1988) A New Metabolic Pathway for meta Ring-Fission Compounds of Biphenyl. *Agric.Biol.Chem.*,52,503-509.

I. Palva (1989) Engineering for Secretion of Proteins by Bacteria. *Soc.Gen.Microbiol. Symp.*,44,255-269.

E.T. Papoutsakis (1984) Equations and Calculations for Fermentions of Butyric Acid Bacteria. *Biotech.Bioeng.*,*26*,174-187.

E.T. Papoutsakis, C.L. Meyer (1985b) Fermentation Equations for Propionic-Acid Bacteria and Production of Assorted Oxychemicals from Various Sugars. *Biotechnol. Bioeng.*,27,67-80.

E.T. Papoutsakis, C.L. Meyer (1985a) Equations and Calculations of Product Yields and Preferred Pathways for Butanediol and Mixed-Acid Fermentations. *Biotechnol. Bioeng.*,27,50-66.

S.W. Peretti, J.E. Bailey (1986) Mechanistically Detailed Model of Cellular Metabolism for Glucose-Limited Growth of *Escherichia coli* B/r-A. *Biotechnol.Bioeng.*, 28,1672-1689.

J. Powlowski, L. Sahlman, V. Shingler (1993) Purification and Properties of the Physically Associated meta-Cleavage Pathwaay Enzymes 4-Hydroxy-2-Ketovalerate Aldolase and Aldehyde Dehydrogenase (Acylating) from Pseudomonas sp. Strain CF600. *J.Bacteriol.*,175,377-385.

C.D. Prater, A.J. Silvestri, J. Wei (1967) On the Structure and Analysis of Complex Systems of

First-Order Chemical Reactions Containing Irreversible Steps. I General Properties. *Chem.Eng.Sci.*,**22**,1587-1606.

A.J. Pratt (1989) B-Lactam Biosynthesis. *Soc.Gen.Microbiol.Symp.*,**44**,163-185.

I. Prigogine (1969) Structure,Dissapation and Life. Theoretical Physics and Biology. pp.23-61 Ed. M.Marois North-Holland Publ.

P.A. Quant (1993) Experimental Aplication of Top-Down Control Analysis to Metabolic Systems. *TIBS*,**18**,26-30.

D. Ramkrishna, S.K. Dhinakar, G.T. Tsao (1987) Are Microbes Optimal Strategists? *Biotechnol.Prog.*,**3**,121-126.

J.L. Ramos, E. Duque, M.I. Ramosgonzalez (1991) Survival In Soils Of An Herbicide-Resistant *Pseudomonas putida* Strain Bearing A Recombinant TOL Plasmid. *Appl.Env.Microbiol.*,**57**,260-266.

J.L. Ramos, C. Michan, F. Rojo, D. Dwyer, K.N. Timmis (1990) Signal-Regulator Interactions. Genetic Analysis of the Effector Binding Site of xylS, the Benzoate-activated Regulator of Pseudomonas TOL Plasmid meta-Cleavage Pathway Operon. *J.Mol.Biol.*,**211**,373-382.

J.L. Ramos, A. Stoltz, W. Reineke, K.N. Timmis (1986) Altered Effector Specificities in Regulators of Gene Expression: TOL Plasmid xylS Mutants and Their Use to Engineer Expansion of the Range of Aromatics Degraded by Bacteria. *Proc.Natl.Acad. Sci.USA.*,**83**,8467-8471.

J.L. Ramos, A. Wasserfallen, K. Rose, K.N. Timmis (1987) Redesigning Metabolic Routes:Manipulation of TOL Plasmid Pathway for Catabolism of Akylbenzoates. *Science*,**235**,593-596.

T.A. Rapoport, R. Heinrich (1975) Mathematical Analysis of Multienzyme Systems: 1 Modeling of the Glycolysis of Human Erythrocytes. *Biosystems*,**7**,120-129.

T.A. Rapoport, R. Heinrich, G. Jacobasch, S. Rapoport (1974) A Linear Steady-State Treatment of Enzymatic Chains: A Mathematical Model of Glycolysis of Human Erythrocytes. *Eur.J.Biochem.*,**42**,107-120.

K.F. Reardon, T.-H. Scheper, J.E.Bailey (1987) Metabolic Pathway Rates and Culture Fluorescence in Batch Fermentations of Clostridium Acetobutylicum. *Biotechnol.Prog.*, **3**,153-167.

C. Reder (1988) Metabolic Control Theory: A Structural Approach. *J.theor.Biol.*, **135**,175-201.

H.E. Reiling, H. Laurila, A. Feichter (1985) Mass Culture of *Escherichia coli:* Medium Development for low and high density cultivation of *Escherichia coli* B/r in Minimal and Complex Media. *J.Biotech.*,**2**,191-206.

A.M. Reiner (1972) Metabolism of Aromatic Compounds in Bacteria. *J.Biol.Chem.*, **247**,4960-4965.

V. Renganathan (1989) Possible Involvement of Toluene-2,3-Dioxygenase in Defluorination of 3-Fluoro-Substituted Benzenes by Toluene-Degrading Pseudomonas sp. Strain T-12. *Appl.Environ.Microbiol.*,**55**,330-334.

W.C. Rheinboldt (1976) On Measures of Ill-Conditioning for Nonlinear Equations. *Math. of Comp.*,**30**,104-111.

J.M. Riol-Cimas, E. Melendez-Hevia (1988) Kinetics of Metabolic Pathways. Transient Response of the

Glycolytic System After Phosphofructokinase Reaction to ADP Input. *Int.J.Biochem.*,**20**,29-33.

J.A. Roels (1980) Application of Macroscopic Principles to Microbial Metabolism. *Biotechnol.Bioeng.*,**22**,2457-2514.

F. Rojo, D.H. Pieper, K.-H. Engesser, H.-J. Knackmuss, K.N. Timmis (1987) Assemblage of Ortho Cleavage Route for Simultaneous Degradation of Chloro- and Methylaromatics. *Science*,**253**,1395-1398.

R.K. Rothmel, A.M. Chakrabarty, A. Berry, A. Darzins (1991) Genetic Systems In Pseudomonas. *Meth.Enzymology*,**204**,485-514.

H. Rozie, H. Kamphuis, W. Somers, J. Visser, F. Rombouts (1987) The Development of an Affinity Adsorbant to Purify Bacterial α-Amylase. *Affin.Chromoatography*, **368**,770-771.

M. Rutgers, K. van Dam, H.V. Westerhoff (1991) Control and Thermodynamics of Microbial Growth: Rational Tools for Bioengineering. *Crit.Rev.Biotechnol.*,**11**,367-395.

J.M. Sala-Trepat, W.C. Evans (1971) The meta-Cleavage of Catechol by Azotobacter Species. *Eur.J.Biochem.*,**20**,400-413.

E. Santero, T.R. Hoover, A.K. North, D.K. Berger, S.C. Porter, S. Kustu (1992) Sigma-Dependent nifH Promoter. *J.Mol.Biol.*,**227**,602-620.

H.M. Sauro, J.R. Small, D.A. Fell (1987) Metabolic Control and its Analysis. *Eur.J. Biochem.*,**165**,215-221.

M.A. Savageau (1969a) Biochemical Systems Analysis: I.Some Mathematical Properties of the Rate Law for the Component Enzymatic Reactions. *J.theor.Biol.*,**25**,365-369.

M.A. Savageau (1970) Biochemical Systems Analysis: III.Dynamic Solutions Using a Power-Law Approximation. *J.theor.Biol.*,**26**,215-226.

M.A. Savageau (1991) Biochemical Systems Theory: Operational Differences Among Variant Representations and their Significance. *J.theor.Biol.*,**151**,509-530.

M.A. Savageau (1969b) Biochemical Systems Analysis: II.The Steady State Solutions for an n-Pool System Using a Power-Law Approximation. *J.theor.Biol.*,**25**,370-379.

M.A. Savageau, A. Sorribas (1989) Constraints Among Molecular and Systemic Properties: Implications for Physiological Genetics. *J.theor.Biol.*,**141**,93-115.

M.A. Savageau, E.O. Voit (1982) Power-Law Approach to Modelling Biological Systems. I. Theory *J.Ferment.Technol.*,**60**,221-228.

M.A. Savageau, E.O. Voit, D.H. Irvine (1987a) Biochemical Systems Theory and Metabolic Control Theory. I. Fundamental Similarities and Differences. *Math.Biosci.*, **86**,127-145.

M.A. Savageau, E.O. Voit, D.H. Irvine (1987b) Biochemical Systems Theory and Metabolic Control Theory. II. The Role of Summation and Connectivity Relationships. *Math.Biosci.*,**86**,147-169.

J.M. Savinell, B.O. Palsson (1992b) Optimal Selection of Metabolic Fluxes for in vivo Measurement. II. Application to Escherichia coli and Hybridoma Cell Metabolism. *J.theor.Biol.*,**155**,215-242.

J.M. Savinell, B.O. Palsson (1992) Optimal selection of Metabolic Fluxes for in vivo Measurement. I. Development of Mathematical Methods. *J.theor.Biol.*,**155**,201-214.

J.M. Savinell, B.O. Palsson (1992a) Network Analysis of Intermediary Metabolism Using Linear Optimization. I. Development of Mathematical Formalism. *J.theor.Biol.*, **154**,421-455.

J.M. Savinell, B.O. Palsson (1992b) Network Analysis of Intermediary Metabolism. II. Interpretation of Hybridoma Cell Metabolism. *J.theor.Biol.*,**154**,455-473.

F.J. Schendel, E.J. Baude, M.C. Flickinger (1989) Determination of Protein Expression and Plasmid Copy Number from Cloned Genes in Escherichia coli by Flow Injection Analysis Using an Enzyme Indicator Vector. *Biotechnol.Bioeng.*,**34**,1023-1036.

P.M. Schlosser, J.E. Bailey (1990) An Integrated Modelling-Experimental Strategy for the Analysis of Metabolic Pathways. *Math.Biosciences*,**100**,87-114.

R. Schuster, H.-G. Holzhutter, G. Jacobasch (1988) Interrelations between Glycolysis and the Monophosphate Shunt in Erythrocytes as Studied on the Basis of a Mathematical Model. *Biosystems*,**22**,19-36.

R. Schuster, G. Jacobasch, H.-G. Holzhutter (1989) Mathematical Modeling of Metabolic Pathways Affected by an Enzyme Deficiency. *Eur.J.Biochem.*,**182**,605-612.

S.A. Selifonov, I.I. Starovoitov (1990) Purification And Characterization Of The Nadh - Acceptor Reductase Comparative-Study Of The Enzymes Of Meta-Cleavage Of The Aromatic Ring In Strains Of The Bacterium Pseudomonas With Plasmid And Chromosomal Genetic Control Of Biphenyl And M-Toluate Catabolism. *Biochem.USSR*, **55**,1616-1624.

A.K. Sen (1990) Topological Analysis of Metabolic Control. *Math.Biosci.*,**102**,191-223.

A. Seressiotis, J.E. Bailey (1988) MPS: An Artificially Intelligent Software System for the Analysis and Synthesis of Metabolic Pathways. *Biotechnol.Bioeng.*,**31**,587-602.

J.P. Shaw, S. Harayama (1992) Component Of Xylene Monooxygenase Encoded By The Tol Plasmid Pww0 Of Pseudomonas-Putida Mt-2. *Eur.J.Biochem.*,**209**,51-61.

J.P. Shaw, F. Schwager, S. Harayama (1992) Substrate Specificity Of Benzyl Alcohol-Dehydrogenase And Benzaldehyde Dehydrogenase Encoded By Tol Plasmid Pww0 - Metabolic And Mechanistic Implications. *Biochem.J.*,**283**,789-794.

M.I. Sinclair, B.W. Holloway (1991) Chromosomal Insertion Of Tol Transposons In Pseudomonas-Aeruginosa Pao. *J.Gen.Microbiol.*,**137**,1111-1120.

V. Shingler, J. Powlowski, U. Marklund (1992) Nucleotide Sequence and Functional Analysis of the Complete Phenol/3,4-Dimethylphenol Catabolic Pathway of Pseudomonas sp. Strain CF600. *J.Bacteriol.*,**174**,711-724.

J. Shu, M.L. Shuler (1989) A Mathematical Model for the Growth of a Single Cell of E.coli on a Glucose/Glutamine/Ammonium Medium. *Biotechnol.Bioeng.*,**33**,1117-1126.

H. Sies (1982) Nicotanamide Nucleotide Compartmentation. Metabolic Compartmentation pp 205-231 Ed. H. Sies.

A.J. Silvestri, C.D. Prater, J. Wei (1968) On the Structure and Analysis of Complex Systems of First-Order Chemical Reactions Containing Irreversible Steps. II Projection Properties of the Characteristic

Vectors. *Chem.Eng.Sci.*,**23**,1191-1200.

P.J. Slininger, L.E. Brebstrator, J.M. Lomont, B.S. Dien, M.R. Oros, M.R. Ladisch, R.J. Bothast (1990) Stoichiometry and Kinetics of Xylose Fermentation by *Pichia stipitis*. *Annals N.Y.Acad.Sci.Biochem.Eng.***VI**,25-40.

J.R. Small, D.A. Fell (1989) The Matrix Method of Metabolic Control Analysis: Its Validity for Complex Pathway Structures. *J.theor.Biol.*,**136**,181-197.

D.J. Smith, M.K.R. Burnham, J.H. Bull, J.E. Hodgson, J.M. Ward, P. Browne, J. Brown, B. Barton et al (1990) β-Lactam Antibiotic Biosynthetic Genes Have Been Conserved in Clusters in Prokaryotes and Eucaryotes. *EMBO J.*,**9**,741-747.

D.J. Smith, M.K.R. Burnham, J. Edwards, A.J. Earl, G. Turner (1990) Cloning and Heterologous Expression of the Penicillin Biosynthetic Gene Cluster from Penicillium chrysogenum. *Biotechnology*,**8**,39-41.

A. Sorribas, M.A. Savageau (1989a) A Comparison of Variant Theories of Intact Biochemical Systems. I. Enzyme-Enzyme Interactions and Biochemical Systems Theory. *Math.Biosci.*,**94**,161-193.

A. Sorribas, M.A. Savageau (1989b) A Comparison of Variant Theories of Intact Biochemical Systems. II. Flux-Oriented and Metabolic Control Theories. *Math.Biosci.*, **94**,195-237.

A. Sorribas, M.A. Savageau (1989c) Strategies for Representing Metabolic Pathways Within Biochemical Systems Theory: Reversible Pathways. *Meth.Biosci.*,**94**,239-269.

J.C. Spain, G.J. Zylstra, C.K. Blake, D.T. Gibson (1989) Monohydroxylation of Phenol and 2,5-Dichlorophenol by Toluene Dioxygenase in *Pseudomonas putida*. *Appl.Environ.Microbiol.*,**55**,2648-2652.

D.E. Steinmeyer, M.L. Shuler (1989) Structured Model for *Saccheromyces cerevisae*. *Chem.Eng.Sci.*,**44**,2017-2030.

G. Stephanopoulos, G. Stephanopoulos (1986) Artificial Intelligence in the Development and Design of Biochemical Processes. *TIBTECH*,**4**,241-249.

G. Stephanopoulos, J.J. Vallino (1991) Network Rigidity and Metabolic engineering in Metabolite Overproduction. *Science*,**252**,1675-1680.

G.M. Stephens, H. Dalton (1988) Kinetics of Benzoate-Induced Loss of the TOL Plasmid from *Pseudomonas putida* MT15 During Growth in Chemostat Culture. *FEMS Microbiol.Letters*,**55**,175-180.

G.M. Stephens, J.M. Sidebotham, N.H. Mann, H. Dalton (1989) Cloning and Expression in *Escherichia coli* of the Toluene Dioxygenase Gene from *Pseudomonas putida* NCIB11767. *FEMS Microbiol.Letters*,**57**,295-300.

K.M. Stone, F.W. Roach, N.F. Thornhill (1991) Dry Weight Measurement of Microbial Biomass and Measurement Variability Analysis.

J.W. Stucki (1982) Thermodynamic Optimizing Principles in Mitochondrial Energy Conversions. Metabolic Compartmentation pp 39-69 Ed. H. Sies

V. Subramanian, T.-N. Liu, W.-K. Yeh, D.T. Gibson (1979) Toluene Dioxygenase: Purification of an

Iron-Sulphate Protein by Affinity Chromatography. *Biochem.Biophys. Res.Commun.*,**91**,1131-1139.

V. Subramanian, T.-N. Liu, W.-K. Yeh, M. Narro, D.T. Gibson (1981) Purfication and Properties of NADH-Ferrodoxin TOL Reductase. *J.Biol.Chem.*,**256**,2723-2730.

M. Suzuki, T. Hayakawa, J.P. Shaw, M. Rekik, S. Harayama (1991) Primary Structure Of Xylene Monooxygenase - Similarities To And Differences From The Alkane Hydroxylation System. *J.Bacteriol.*,**173**,1690-1695.

K. Taira, J. Hirose, S. Hayashida, K. Furukawa (1992) Analysis of bph Operon from the Polychlorinated Biphenyl-Degrading Strain of *Pseudomonas pseudoalcaligenes* KF707. *J.Biol.Chem.*,**267**,4844-4853.

K. Taira, K. Nakagawa, S. Nishikawa, K. Furukawa (1991) Construction of a Novel RNA-Transcript-Trimming Plasmid Which Can be Used Both in vitro in place of Run-Off and (G)-Free Transcriptions and in vivo as Multi-Sequences trannscription Vectors. *Nucleic Acids Res.*,*19*,5125-5130.

P. Thibault, C. Paris, S. Pleasance (1991) Analysis of Peptides and Proteins by Capillary Electrophoresis/Mass Spectrometry Using Acidic Buffers and Coated Capillaries. *Rap.Comm.Mass Spec.*,**5**,484-490.

A.W. Thomas, J.H. Slater, A.J. Weightman (1992) The Dehalogenase Gene dehI From Pseudomonas Putida Pp3 Is Carried On An Unusual Mobile Genetic Element Designated deh. *J.Bacteriol.*,**174**,1932-1940.

A.W. Thomas, A.W. Topping, J.H. Slater, A.J. Weightman (1992) Localization And Functional-Analysis Of Structural And Regulatory Dehalogenase Genes Carried On Deh From Pseudomonas Putida Pp3. *J.Bacteriol.*,**174**,1941-1947.

J.A. Thomas, M.A. Valvano (1992) TolQ Is Required For Cloacin Df13 Susceptibility In Escherichia-Coli Expressing The Aerobactin Cloacin Df13 Receptor Iuta. *FEMS Microbiol.Lett.*,**91**,107-111.

N.V. Torres, F. Mateo, E. Melendez-Hevia, H. Kacser (1986) Kinetics of Metabolic Pathways. *BiochemJ.*,**234**,169-174.

N.V. Torres, F. Mateo, J. Sicilia, E. Melendez-Havia (1988) Distribution of the Flux Control in Convergent Metabolic Pathways: Theory and Application to Experimental and Simulated Systems. *Int.J.Biochem.*,**20**,161-165.

L.B. Tsai, M. Mann, F. Morris, C. Rotgers, D. Fenton (1987) The Effect of Organic Nitrogen and Glucose on the Production of Recombinant Human Insulin-Like Growth Factor in High Cell Density Escherichia coli Fermentations. *J.Ind.Microbiol.*,**2**,181-187.

S.P. Tsai, Y.H. Lee (1988b) Application of Gibbs' Rule and a Simple Pathway Method to Microbial Stoichiometry. *Biotech.Prog.*,**4**,82-88.

S.P. Tsai, Y.H. Lee (1988a) Application of Metabolic Pathway to Statistical Analysis of Bioreactor Measurement Data. *Biotech.Bioeng.*,**32**,713-715.

J.H.A.A. Uitzetter, C.J. Bos, J. Visser (1986) Characterization of Aspergillus nidulans Mutants in Carbon Metabolism Isolated After D-Galacturonate Enrichment. *J.Gen. Microbiol.*,**132**,1167-1172.

J.R. Vandermeer, W.M. Devos, S. Harayama, A.J.B. Zehnder (1992) Molecular Mechanisms Of Genetic

Adaptation To Xenobiotic Compounds. *Microbiol.Rev.*,**56**, 677-694.

J.R. Vandermeer, A.J.B. Zehnder, W.M. Devos (1991) Identification Of A Novel Composite Transposable Element, Tn5280, Carrying Chlorobenzene Dioxygenase Genes Of Pseudomonas Sp Strain P51. *J.Bacteriol.*,**173**,7077-7083.

S.E. Vecht, M.W. Platt, Z. Er-El, I. Goldberg (1988) The Growth of Pseudomonas putida on m-Toluic Acid and on Toluene in Batch and in Chemostat Cultures. *Appl.Microbiol.Biotechnol.*,**27**,587-592.

L.C. Vining, S. Shapiro, K. Madduri, C. Stuttard (1990) Biosynthesis and the Control of β-Lactam Antibiotics: the Early Steps in the "Classical" Tripeptide Pathway. *Biotech.Adv.*,**8**,159-183.

J. Visser, R.van Rooijen, C. Dijkema, K. Swart, H.M. Sealy-Lewis (1988) Glycerol Uptake Mutants of the Hyphal Fungus *Aspergillus nidulans*. *J.Gen.Microbiol.*,**134**, 655-659.

E.O. Voit (1992) Optimization in Integrated Biochemical Systems. *Biotechnol.Bioeng.*, **40**,572-582.

E.O. Voit, M.A. Savageau (1987) Accuracy of Alternative Representations for Integrated Biochemical Systems. *Biochemistry*,**26**,6869-6880.

E.O. Voit, M.A. Savageau (1982) Power-Law Approach to Modelling Biological Systems. II. Application to Ethanol Production. *J.Ferment.Technol.*,**60**,229-232.

L.P. Wackett, S.R. Householder (1989) Toxicity of Trichloroethylene to Pseudomonas putida F1 is Mediated by Toluene Dioxygenase. *Appl.Environ.Microbiol.*,**55**,2723-2725.

L.P. Wackett, L.D. Kwart, D.T. Gibson (1988) Benzylic Monooxygenation Catalyzed by Toluene Dioxygenase from *Pseudomonas putida*. *Biochemistry*,**27**,1360-1367.

K. Walsh, D.E. Koshland (1985) Characterization of Rate-Controlling Steps in vivo by Use of an Adjustable Expression Vector. *Proc.Natl.Acad.Sci.USA.*,**82**,3577-3581.

D.I.C. Wang, C.L. Cooney, A.L. Dewain, P. Dunnill, A.E. Humphrey, M.D. Lilly (19) Fermentation and Enzyme Technology.

A. Wasserfallen, M. Rekik, S. Harayama (1991) A Pseudomonas putida Strain Able To Degrade M-Toluate In The Presence Of 3-Chlorocatechol. *Bio/Technology*,**9**, 296-298.

G.R. Welch, T. Keleti (1990) Metabolic Control Analysis for Heterogenous Enzyme Systems. Structural and Organizational Aspects of Metabolic Regulation.

H.V. Westerhoff, M.A. Aon, K. van Dam, S. Cartessa, D. Kahn, M. van Workum (1990) Dynamic and Hierarchical Coupling. *Biochim.Biophys.Acta.*,**1018**,142-146.

H.V. Westerhoff, J.C. Arents (1984) Two (Completely) Rate-Limiting Steps in One Metabolic Pathway? The Resolution of a Paradox... *Biosci.Rep.*,**4**,23-31.

H.V. Westerhoff, Y.-D. Chen (1984) How Do Enzyme Activities Control Metabolite Concentrations? An Additional Theorem in the Theory of Metabolic Control. *Eur.J.Biochem.*,**142**,425-430.

H.V. Westerhoff, K. Van Dam (1987) Thermodynamics and Control of Biological Free-Energy Transduction.

H.V. Westerhoff, A.K. Groen, R.J.A. Wanders (1984) Modern Theories of Metabolic Control and Their Applications. *Biosci.Rep.*,**4**,1-22.

H.V. Westerhoff, D.B. Kell (1987) Matrix Method for Determinihng Steps Most Rate-Limiting to Metabolic Fluxes in Biotechnological Processes. *Biotechnol.Bioeng.*, **30**,101-107.

H.V. Westerhoff, J.S. Lolkema, R. Otto, K.J. Hellingwerf (1982) Thermodynamics of Growth: Nonequilibrium Thermodynamics of Bacterial Growth, The Phenomenological and the Mosaic Approach. *Biochim.Biophys.Acta.*,**683**,181-220.

Z.C. Wey-Bang, A. Constantinides (1988) Modeling, Optimization and Computer Control of the Cephalosporin C Fermentation Process. *Biotechnol.Bioeng.*,**32**,277-288.

G.M. Whited, W.R. McCombie, L.D. Kwart, D.T. Gibson (1986) Identification of cis-Diols as Intermediates in the Oxidation of Aromatic Acids by a Strain of *Pseudomonas putida* That Contains a TOL Plasmid. *J.Bacteriol.*,**166**,1028-1039.

C. Winstanley, J.A.W. Morgan, R.W. Pickup, J.R. Saunders (1991) Use Of A XylE Marker Gene To Monitor Survival Of Recombinant *Pseudomonas putida* Populations In Lake Water By Culture On Nonselective Media. *Appl.Env.Microbiol.*,**57**,1905-1913.

B. Witholt, O. Favre-Bulle, R. Lageveen, J. Kingma, J.B. van Bielen, H. Marvin, H. Preusting (1989) Synthesis of Apolar Organic Compounds by Pseudomonas spp. and *Escherichia coli* in Two-Liquid Phase Fermentations.

C.B.F. Witteveen, R. Busink, P.van de Vondervoort, C. Dijkema, K. Swart, J. Visser (1989) L-Arabinose and D-Xylose Catabolism in *Aspergillus niger. J.Gen.Microbiol.*, **135**,2163-2171.

M.G. Wubbolts, K.N. Timmis (1990) Biotransformation of Substituted Benzoates to the Corresponding cis-Diols by an Engineered Strain of Pseudomonas oleovorans Producing the TOL Plasmid-Specified Enzyme Toluate-1,2-Dioxegenase. *Appl.Env.Microbiol.*, **56**,569-571.

J. Zeyer, P.R. Lehrbach, K.N. Timmis (1985b) Use of Cloned Genes of Pseudomonas TOL Plasmid to Effect Biotransformation of Benzoates to cis-Dihydrodiols and Catechols by *Escherichia coli* Cells. *Appl.Env.Microbiol.*,**50**,1409-14.

L. Zhou, K.N. Timmis, J.L. Ramos (1990) Mutations Leading to Constitutive Expresion from the TOL Plasmid meta-Cleavage Pathway Operon Are Located at the C-Terminal End of the Positive Regulator Protein xylS. *J.Bacteriol.*,**172**,3707-3710.

G.J. Zylstra, D.T. Gibson (1989a) Toluene Degradation by *Pseudomonas putida* F1: Nucleotide Sequence of the todC1C2BADE Genes and Their Expression in *Escherichia coli. J.Biol.Chem.*,**264**,14940-14946.

G.J. Zylstra, L.P. Wackett, D.T. Gibson (1989b) Trichloroethylene Degradation by Escherichia coli Containing the Cloned *Pseudomonas putida* F1 Toluene Dioxygenase Genes. *Appl.Environ.Microbiol.*,**55**,3162-3166.